

E.T.S. de Ingeniería Industrial,
Informática y de Telecomunicación

DISEÑO, GRABACIÓN Y REPRODUCCIÓN DE PAISAJES SONOROS MEDIANTE AMBISONICS



Grado en Ingeniería
en Tecnologías de Telecomunicación

Trabajo Fin de Grado

Autor: Maitane Vicuña Zubiria

Director: Ricardo San Martín Murugarren

Pamplona, 15 de junio de 2018

AGRADECIMIENTOS

Se agradece la ayuda a los integrantes del Departamento de Acústica, en especial al profesor Ricardo San Martín, el cual, ha estado presente a lo largo del proyecto. Por otro lado, se agradece la colaboración a las personas que ayudaron en la grabación del Camino Fluvial y en la ayuda con la creación de la aplicación. Finalmente, se agradece a los participantes que realizaron los test para la valoración del proyecto.

RESUMEN

El trabajo de fin de grado consiste en la implementación de la tecnología Ambisonics como solución de audio espacial para la recreación de paisajes sonoros. Se han evaluado sus prestaciones tanto para la reproducción de audio en sistemas de altavoces, como en sistemas de escucha binaural.

Durante el proyecto se profundiza en la síntesis de paisajes sonoros ficticios y la grabación de escenas reales, y se ha analizado su posterior reproducción utilizando los diferentes formatos existentes (Formato-A, Formato-B FuMa y Ambix). Se han propuesto diferentes aplicaciones y, finalmente, se ha realizado una valoración mediante test subjetivos.

ABSTRACT

The final degree project consists of the implementation of the Ambisonics technology as a spatial audio solution for the recreation of sound landscapes. Its features have been evaluated both for audio reproduction in landspeaker systems and in binaural listening systems.

During the project, the synthesis of fictitious soundscapes and the recording of real scenes is studied, and its subsequent reproduction has been analyzed using the different existing formats (Format-A, Format-B FuMa and Ambix). Different applications have been proposed and, finally, an assessment has been made by subjective tests.

LISTA DE PALABRAS CLAVE

- Ambisonics: Solución tecnológica de audio espacial
- FuMa / Ambix: Dos convenciones del formato B
- Reaper: Programa de producción de audio
- DAW (Digital Audio Workstation): Estación de trabajo de audio digital
- FB360 Spatial Workstation
- Binaural: Recreación de sonido 3D mediante 2 canales

ÍNDICE

Agradecimientos	2
Resumen/ Abstract	3
Lista de palabras clave	4
1 INTRODUCCIÓN AMBISONICS	6
1.1 Técnicas de grabación	7
1.1.1 Microfonía	7
1.1.2 Formatos A y B.....	9
1.2 Formatos de codificación	11
1.2.1 Compatibilidad estéreo	12
1.3 Formatos de decodificación	13
1.4 Técnicas de reproducción	14
1.5 Conversión a binaural	14
2 APLICACIONES	15
2.1 Introducción de audio espacial en videos 360°	15
2.2 Dispositivo experimental y software utilizado	18
2.3 Diseño de paisajes sonoros	19
2.3.1 Juego de colores	19
I. Introducción	19
II. Escenario	19
III. Procedimiento	21
IV. Resultados	29
2.3.2 Juego selva	30
I. Introducción	30
II. Escenario	30
III. Procedimiento	31
IV. Resultados	36
2.4 Grabación y reproducción de paisajes sonoros	37
2.4.1 Aplicación Camino Fluvial	37
I. Introducción	37
II. Escenario	37
III. Procedimiento	38
2.5 Evaluación mediante test subjetivos de paisajes sonoros	44
I. Introducción	44
II. Procedimiento	45
III. Resultados	47
3 CONCLUSIONES	48
4 BIBLIOGRAFÍA	50
5 ANEXOS	51

1. INTRODUCCIÓN AMBISONICS

El sistema Ambisonic de sonido envolvente, es una solución tecnológica que permite grabar, manipular y reproducir audio en sistemas de altavoces prácticos para que los oyentes puedan percibir los sonidos ubicados en el espacio tridimensional. Esto puede ocurrir a lo largo de un [1] escenario horizontal de 360 grados (sistemas pantofónicos) o a lo largo de una esfera completa (sistemas perifónicos).

Ambisonics se desarrolló a principios de la década de 1970, por un grupo de investigadores británicos, bajo los auspicios de la British NRDC. El principal impulsor fue [2] Michael Gerzon. El nombre surgió a partir de raíces latinas que denotan sonido envolvente; el verbo ambire (ir alrededor, rodear) y el sustantivo sonus (sonido).

El enfoque básico es tratar una escena de audio como una esfera de sonido de 360 grados proveniente de diferentes direcciones alrededor de un punto central. El punto central es donde se coloca el micrófono durante la grabación, o donde se encuentra el "punto óptimo" del oyente durante la reproducción.

A pesar de su sólida base técnica, Ambisonics no fue adoptado comercialmente hasta hace poco con el desarrollo de la industria de realidad virtual que requiere soluciones de audio de 360 °.

La tecnología Ambisonics está distribuida en 4 fases [1]:

- Grabación con técnicas microfónicas
- Codificación en formatos específicos
- Decodificación en n canales
- Reproducción a través de m altavoces

1.1 TÉCNICAS DE GRABACIÓN

La teoría matemática de las señales usadas por Ambisonics, es conocida como la [3] teoría de los armónicos esféricos, donde se puede reconstruir una señal por medio de la suma de componentes individuales. La cantidad de armónicos que se empleen se conoce como el orden de la teoría Ambisonics. Se entiende como orden cero W la función omnidireccional. El primer orden incluye el orden cero y las señales X , Y , Z , el segundo; además de estos cuatro incluye cinco armónicos más y así sucesivamente. En este proyecto, trabajaremos con el primer orden.

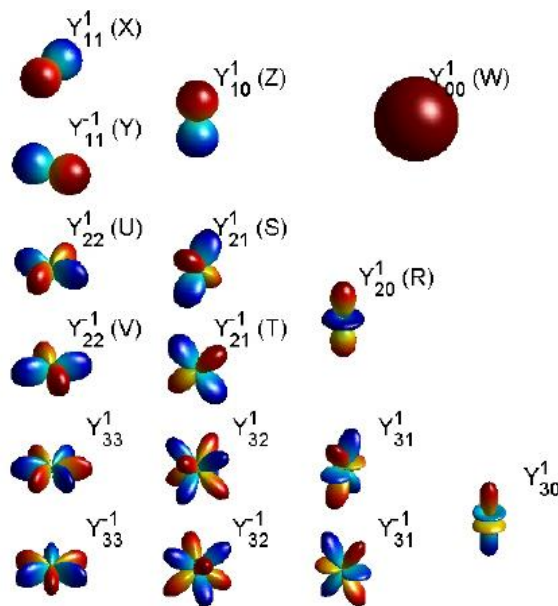


Figura 1. Armónicos esféricos utilizados para reconstruir las señales mediante Ambisonics

1.1.1 MICROFONÍA

Teniendo en cuenta la explicación anterior, existen dos técnicas microfónicas con las cuales obtener un registro adecuado para la codificación propia de Ambisonics [1]:

- **Técnica clásica:** Utiliza una composición de 4 micrófonos coincidentes. Por un lado, se coloca un micrófono omnidireccional para captar el ambiente y ,por otro lado, se añaden tres micrófonos bidireccionales, cada uno orientado en cada eje de coordenadas.

Con esta configuración, se consigue un registro de las variaciones de fase e intensidad de la señal acústica en los tres ejes del espacio con cada uno de los

micrófonos bidireccionales, y un registro global gracias al micrófono omnidireccional. De esta manera se puede capturar toda la información y variaciones del espacio tridimensional, consiguiendo con esto codificar la información acústica en formato-B, de forma que permita luego decodificarla a un número arbitrario de canales y diversas configuraciones de distribución de altavoces.

Una condición indispensable para la efectividad de esta técnica es que las capsulas de los micrófonos deben ser absolutamente coincidentes, algo muy complicado y que puede afectar al resultado final si se utilizan micrófonos independientes. Además, se producen errores en la forma del patrón polar o en la relación entre la sensibilidad de campo directo a campo difuso de los micrófonos.

Este problema se ha solucionado en gran medida con micrófonos creados específicamente para la técnica siguiente.

- Técnica Ambisonic: Utiliza un único micrófono que contiene una matriz tetraédrica de cápsulas de micrófono nominalmente cardioides. Las señales de la cápsula se procesan de manera que den cuatro señales de salida, en formato-B, que son proporcionales a la presión y el vector de velocidad de partícula tridimensional en el centro de la matriz. Algunos de los modelos que han salido al mercado son; Soundfield, TetraMic y Sennheiser Ambeo. Este último será el que utilicemos en el proyecto.

Sennheiser Ambeo VR 3D

Este micrófono Ambisonico captura audio esférico inmersivo [4] mediante 4 capsulas KE14 acomodadas en forma de tetraedro. Este diseño permite capturar esféricamente el sonido proveniente de todas las direcciones a partir de un solo punto en el espacio. Las cápsulas proporcionan 4 canales en formato-A, que tienen que convertirse en un nuevo set de 4 canales en formato-B. Para ello, Sennheiser incluye software convertidor de formato A-B.



Figura 2. Imágenes del micrófono AMBEO VR 3D

1.1.2 FORMATOS A Y B

Las configuraciones microfónicas existentes dan como resultado cuatro señales que están relacionadas entre sí de cierta manera, es por ello que de cada una de las técnicas microfónicas se obtiene un formato diferente:

- A-Format: Cuando se realiza una grabación con los micrófonos Ambisonics, se obtienen cuatro señales en el llamado formato-A. Debido a que el formato en el que trabaja Ambisonics es el formato-B, es necesario realizar una conversión de formato A-B para poder trabajar con esas señales.

Este tipo de micrófonos utiliza [1] cuatro capsulas colocadas en forma de tetraedro. Empleando el sistema de referencia cartesiano antropométrico (ISO 2631), se etiquetan las señales de las cuatro cápsulas de la siguiente manera:

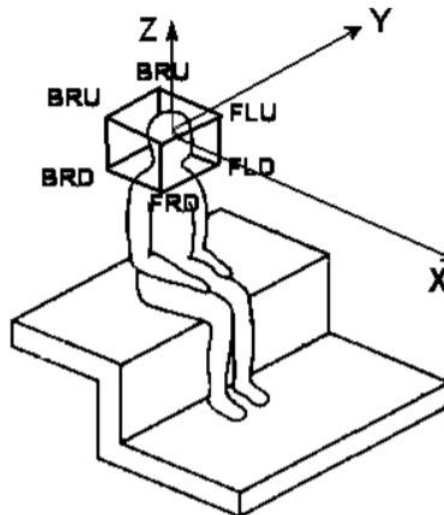


Figura 3. Posicionamiento de las cápsulas en cada eje

Las cápsulas corresponden a cuatro de los ocho vértices del cubo que rodea la cabeza del oyente virtual. Este cubo tiene caras paralelas al sistema de referencia cartesiano.

Cada vértice del cubo imaginario se define por la [1] intersección de tres caras. Así, por ejemplo, un micrófono dirigido hacia adelante, izquierda y arriba será etiquetado FLU (Front-Left-Up). La selección habitual de las posiciones suele ser: FLU-FRD-BLD-BRU

Si tomáramos únicamente el plano horizontal, las señales se ordenan de la siguiente forma:

FL: Front-Left
FR: Front-Right
BL: Back-Left
BR: Back-Right

A continuación, se graban los 4 canales anteriores procedentes de las capsulas procesados adecuadamente sobre el formato-B, sobre los canales W, X, Y y Z.

Las 4 nuevas señales se definirán matemáticamente como las señales de 4 micrófonos coincidentes. Como se ha especificado anteriormente, la señal W corresponderá a un micrófono omnidireccional, mientras que X, Y y Z serán micrófonos de velocidad, con un patrón de directividad que se asemeja a una figura de 8.

Los fundamentos de conversión de A-Format a B-Format se realizan a través de operaciones matriciales de las señales recogidas de las capsulas. Las formulas matriciales serán las siguientes [1]:

$$W' = FLU + FRD + BLD + BRU$$

$$X' = FLU + FRD - BLD - BRU$$

$$Y' = FLU - FRD + BLD - BRU$$

$$Z' = FLU - FRD - BLD + BRU$$

Las señales con apostrofe, no son las señales de salida equivalentes al estándar de Formato-B, ya que, requieren operaciones de filtrado. Estas operaciones de filtrado están regidas por [1] dos filtros distintos, aplicando cada uno de ellos en función del canal de entrada.

- Filtro Fw: aplicado a W' daría como resultado W.

$$F_W = \frac{1 + \frac{j\omega r}{c} - \frac{1}{3}\left(\frac{\omega r}{c}\right)^2}{1 + \frac{1}{3}\left(\frac{j\omega r}{c}\right)}$$

r= distancia de cada capsula al centro del tetraedro en metros
w= frecuencia angular en rad/s ($w=2\pi f$)

c = velocidad del sonido en m/s ($c=340\text{m/s}$)

- Filtro F_{xyz} : aplicado a los canales X' , Y' y Z' daría como resultado X , Y y Z respectivamente.

$$F_{XYZ} = \sqrt{6} \frac{1 + \frac{1}{3} \left(\frac{j\omega r}{c} \right) - \frac{1}{3} \left(\frac{\omega r}{c} \right)^2}{1 + \frac{1}{3} \left(\frac{j\omega r}{c} \right)}$$

r = distancia de cada capsula al centro del tetraedro en metros

w = frecuencia angular en rad/s ($w=2\pi f$)

c = velocidad del sonido en m/s ($c=340\text{m/s}$)

Estos filtros mantienen la ecualización ganacial, pero afectan ligeramente a la fase.

- **B-Format:** Corresponde a las cuatro señales que se obtienen con la configuración microfónica de la técnica clásica. Es la señal estándar en Ambisonics y se identifica cada uno de los canales mediante las siglas W , X , Y y Z .

Existen dos convenciones dentro del estándar de formato B de Ambisonics: AmbiX y FuMa. Son bastante similares, pero no intercambiables: se diferencian por la secuencia en que se organizan los cuatro canales, con AmbiX, por ejemplo, dispuesto WYZX en lugar de WXYZ. En el proyecto se utilizarán ambas convecciones, según la necesidad de cada prueba.

1.2 FORMATOS DE CODIFICACIÓN

Los sonidos posicionados en formato B se colocan conceptualmente en la superficie de una esfera "unidad" o dentro de ella. Si el radio máximo del campo de sonido es 1, los sonidos movidos fuera de esta esfera (con un radio mayor que 1), no se decodificarán correctamente y tenderán a ubicarse hacia el altavoz más cercano. Por ello, las coordenadas de la fuente de sonido deben obedecer la siguiente regla [2]:

$$(x^2 + y^2 + z^2) \leq 1$$

donde x es la distancia a lo largo del eje X (frente-atrás), y es la distancia a lo largo del eje Y (izquierda-derecha), y z es la distancia a lo largo del eje Z (arriba-abajo).

Cuando se coloca una señal monofónica en la superficie de la esfera, sus coordenadas con referencia al centro serán [2]:

$$\begin{aligned}x &= \cos A \times \cos B, \\y &= \sin A \times \cos B, \text{ and} \\z &= \sin B.\end{aligned}$$

Donde A es el ángulo horizontal y B el ángulo vertical (en sentido contrario a las agujas del reloj). Estas coordenadas se usan como multiplicadores para producir las señales de salida de formato-B (X, Y, Z y W) así,

$$\begin{aligned}X &= \text{input signal} \times \cos A \times \cos B, \\Y &= \text{input signal} \times \sin A \times \cos B, \\Z &= \text{input signal} \times \sin B, \text{ and} \\W &= \text{input signal} \times 0.707.\end{aligned}$$

El multiplicador 0.707 en W proporciona una distribución más uniforme de los niveles de señal dentro de los cuatro canales, y es el resultado de consideraciones de ingeniería.

1.2.1 COMPATIBILIDAD ESTÉREO (UHF FORMAT)

Las señales Ambisonic de formato B no son directamente compatibles con estéreo, aunque es posible generar una respuesta exactamente equivalente a la de un par de micrófonos cruzados (Farrah 1979). Sin embargo, hay un sistema británico de dos canales (conocido como UHF) que permite que la mayoría de la información horizontal de las señales W, X e Y de formato B se codifiquen en matriz para formar una señal estéreo. Además, mediante un decodificador UHF adecuado se puede conseguir cierta especialización sonora al reproducirlo por este sistema.

Este método de codificación se usa para reproducir grabaciones, cintas y emisiones radiales estéreo compatibles con el sistema Ambisonics. Para ello, las señales W, X e Y son codificadas en dos canales usando la siguiente transformada [5]:

$$\text{Left} = (0.0928 + 0.255j)X + (0.4699 - 0.171j)W + (0.3277)Y$$

$$\text{Right} = (0.0928 - 0.255j)X + (0.4699 + 0.171j)W - (0.3277)Y$$

Esto sería relativamente simple si no fuera por la parte imaginaria de la ecuación. Lo que esto indica es que esa señal posee un desplazamiento en fase de 90 grados a lo largo de la banda completa de audio respecto a su versión normal. Para hacer esto, cada una de las tres señales deberá pasar a través de desplazadores de fase de 90 grados de banda ancha para codificar y decodificar.

1.3 FORMATOS DE DECODIFICACIÓN

La decodificación de señales codificadas de forma Ambisónica puede parecer complicada. La complejidad aparece en la optimización de decodificadores para sistemas británicos (con un número limitado de altavoces y una pequeña área de escucha) que utilizan técnicas psico acústicas, pero estos no son productivos en sistemas utilizados para cubrir grandes áreas (Malham 1992).

En Ambisonics, cada altavoz individual recibe una combinación de señales de formato B correspondientes a su posición con respecto al centro de la matriz. Por ejemplo, para una [2] matriz cuadrada (horizontal solamente) de cuatro altavoces, dispuestos en el frente izquierdo (LF), el frente derecho (RF), el lado izquierdo atrás (LB) y el derecho atrás (RB), las señales son:

$$\begin{aligned} LF &= W + 0.707(X + Y), \\ RF &= W + 0.707(X - Y), \\ LB &= W + 0.707(-X + Y), \text{ and} \\ RB &= W + 0.707(-X - Y). \end{aligned}$$

Para una matriz cúbica, las señales para las cuatro esquinas planas en los planos "arriba" (U) y "abajo" (D) son [2]:

$$\begin{aligned} LFU &= W + 0.707(X + Y + Z), \\ RFU &= W + 0.707(X - Y + Z), \\ LBU &= W + 0.707(-X + Y + Z), \\ RBU &= W + 0.707(-X - Y + Z), \\ LFD &= W + 0.707(X + Y - Z), \\ RFD &= W + 0.707(X - Y - Z), \\ LBD &= W + 0.707(-X + Y - Z), \text{ and} \\ RBD &= W + 0.707(-X - Y - Z). \end{aligned}$$

El factor de directividad de 0.707 en las ecuaciones anteriores da como resultado una respuesta direccional de origen cardioide para cada altavoz. Esto es óptimo para posiciones de escucha cercanas a los altavoces o fuera de la matriz de altavoces (Malham 1992). Cuando el área de audición prevista es significativamente más pequeña que la matriz de altavoces, se puede emplear una forma más hipercardioide aumentando el factor de directividad, lo que da como resultado una imagen mejorada para los oyentes localizados centralmente.

1.4 TECNICAS DE REPRODUCCIÓN

Los sistemas que utilizan las señales de Formato-B para transportar la información grabada requieren de cuatro canales para una codificación completa de los sonidos hasta la precisión que se puede obtener con micrófonos de primer orden. El número mínimo de altavoces requeridos para la reproducción es:

- Sistemas pantofónicos: cuatro altavoces para localizar las fuentes originales en el plano horizontal.
- Sistemas perifónicos: ocho altavoces para localizar las fuentes originales tanto en el plano horizontal como vertical.

No hay necesidad de considerar los detalles reales del sistema de reproducción cuando se hace una grabación o síntesis, si se siguen las especificaciones del formato B. Ambisonics es el único sistema de sonido envolvente que emplea este tipo de procesamiento psicoacústico para optimizar las señales anteriores a la reproducción sobre las posiciones de los altavoces determinados por el usuario.

Finalmente, no existe una configuración concreta para la reproducción de grabaciones realizadas con Ambisonics, ya que será el propio usuario el que decidirá la configuración de los altavoces en base a las [1] siguientes limitaciones:

- Los altavoces se situarán en la superficie de una esfera en cuyo centro se situará el oyente.
- El número de altavoces necesarios debe cumplir $L \geq (N + 1)^2$, donde L es el número de altavoces y N el orden de codificación.

1.5 CONVERSIÓN A BINAURAL

La industria de realidad virtual ha adoptado Ambisonics como estándar para ofrecer audio 3D en videos, juegos y experiencias de realidad virtual. En ocasiones el usuario final experimenta el audio a través de auriculares. Para ello, se utilizará la tecnología de audio binaural.

La forma más común de procesar Ambisonics para la reproducción espacial binaural en auriculares, es [6] decodificar los canales Ambisonics para una determinada matriz de altavoces, pero luego, en lugar de enviarlos a altavoces reales, los canales se envían a un procesador binaural que prácticamente los ubica

en la dirección que el hablante real lo habría transmitido. Así, finalmente el oyente experimentara con los auriculares un campo de sonido esférico inmersivo.

En el proyecto se va a utilizar el complemento FB360 Spatial Workstation para la conversión a audio binaural.

2. APLICACIONES

2.1 INTRODUCCIÓN DE AUDIO ESPACIAL EN VIDEOS 360°

Para la realización de las siguientes aplicaciones, ha sido necesario estudiar la forma de crear videos 360° introduciendo audio espacial Ambisonics. Con el objetivo de que el oyente pueda sumergirse en el contexto de una escena real, mediante un paisaje sonoro inmersivo.

Se han evaluado diferentes programas que resuelven los dos principales problemas de incluir audio espacial:

- Cómo hacer que un archivo de audio suene en la misma dirección proveniente de la ubicación real, en la escena ficticia. (Espacializar)
- Como conseguir que cualquier audio espacial (Formato B p.e) o archivo de audio no espacial sea interactivo para un espectador.

Los programas más significativos son:

➤ **Adobe Premier Pro**

En la siguiente aplicación se incluyen diversas funcionalidades, entre ellas:

- Edición de video de realidad virtual envolvente, pudiendo editar RV del mismo modo en el que el espectador lo ve, mediante gafas de RV o realizando la edición mediante el teclado.
- Edición de audio envolvente para contenido 360°/RV. Se da la posibilidad de editar el audio en función de la orientación y después se puede exportar como audio Ambisonics en plataformas como Youtube y Facebook.
- Efectos envolventes: Se añaden efectos a videos 360°/RV sin crear distorsiones no deseadas en los polos (desenfoque, resplandor etc.)

-Transiciones envolventes: Se pueden introducir transiciones de movimiento como barridos irisados, de degradado y zooms.

El mayor inconveniente es que se trata de un software de pago, por lo tanto, esta opción ha sido descartada.

➤ **G'Audio Works**

El software aprovecha la mezcla basada en objetos para la colocación detallada del sonido en un video 3D. Además, se incluye un codificador compatible tanto con Mac como con Windows.

El problema principal de G'audio Works radica en que las funciones generales del programa solo están disponibles para dispositivos Mac, por ello, ha sido desplazada de las opciones para utilizar en el proyecto.

➤ **Steam Audio**

El siguiente programa incorpora rotación y seguimiento posicional para RV. Se da la posibilidad de crear una representación binaural basada en HRTF y se puede renderizar audio basado en Ambisonics de 1er orden o superior. Se incorporan controles que permiten usar la geometría de escena existente para ocluir y reflejar sonidos.

El principal inconveniente de dicho programa se debe a que solo recoge videos de RV, no trabaja con videos 360° de paisajes sonoros reales, por ello, también ha sido descartada.

➤ **Avid Pro Tools HD**

El software Avid Pro Tools HD incluye soporte para audio Ambisonic y RV/360. Además, realiza edición de video avanzada con hasta 64 pistas de video. Como desventaja, se puede mencionar que se trata de un programa de pago, por ello, se ha decidido descartar para este proyecto.

➤ **Ambisonic Toolkit**

Es un plugin de audio espacial gratuito para [7] Reaper y Super Collider, que reúne una serie de herramientas para trabajar con Ambisonic Surround Sound. Facilita la composición espacial más allá de la colocación de sonidos en una escena de sonido.

-Reaper: Es un conjunto de plugins para la composición espacial, diseño de sonido y desarrollo de contenido de RV en la estación de audio digital (DAW). Los complementos funcionan en todas las plataformas compatibles con Reaper. Dichos complementos son: Codificadores, transformadores, decodificadores y otras utilidades.

-Super collider: Es un lenguaje de programación para síntesis de audio en tiempo real y composición algorítmica.

Son compatibles con Mac, Windows y Linux. Para este proyecto, se ha decidido utilizar los plugins de ATK para Reaper, por su gran variedad de funciones útiles y su amplia compatibilidad.

➤ **360pan suite with Reaper**

Esta suite es un conjunto de complementos para Windows y Mac diseñado para ofrecer mezclas Ambisonics para audio inmersivo: audio de todas las direcciones, destinado a la reproducción de auriculares, que se mantiene donde esta incluso cuando gira la cabeza.

Se puede hacer todo el paneo, distanciamiento e incluso mezcla desde la ventana de video de su DAW con Reaper o Pro Tools HD. Pero su mayor inconveniente es que el programa es de pago, y solo incluye funciones para la reproducción binaural.

➤ **FB360 spatial Workstation with Reaper**

Se trata de uno de los más completos plugins y es gratuito, por ello, en el proyecto se van a realizar pruebas con esta suite. La suite consta de complementos para Digital Audio Workstations (DAW), un reproductor de video VR y un motor multiplataforma que facilita el diseño y la entrega de contenido para proyectos de video en RV.

Se va a utilizar Ambisonics para describir un campo de sonido 3D completo. Para ello, se emplearán varias pistas de audio interdependientes que pueden decodificarse en otros formatos como por ejemplo binaural, lo que significa que solo se necesita hacer una grabación Ambisonic para múltiples casos de uso diferentes.

Además, una mezcla Ambisonic se puede decodificar dinámicamente para incorporar la rotación de la cabeza del oyente.

Por último, FB360 se puede emplear usando Pro Tools HD o Reaper. Para este proyecto se va a utilizar como ayuda para el programa Reaper.

2.2 DISPOSITIVO EXPERIMENTAL Y SOFTWARE UTILIZADO

Infraestructura Hardware

- Micrófono Sennheiser Ambeo VR 3D
- Trípode para micrófono
- Zoom F4
- 8 altavoces Neumann KH120A
- Ordenador portátil Lenovo Miix
- Interfaz de audio Focusrite 18i20
- Gafas 3D Xiaomi Play 2VR

Infraestructura Software

- **Focusrite Control**

Se trata de un programa que funciona como mezclador y trata de enrutar las señales de entrada y salida entre el ordenador y el interfaz de audio.

- **Reaper Digital Audio Workstation v. 5**

Reaper es el programa principal con el que trabajaremos. Se trata de una aplicación de producción de audio que ofrece un conjunto de herramientas de grabación, edición, procesamiento, mezcla y masterización multipista. Además, admite una amplia gama de hardware, formatos digitales y complementos.

- **The Ambisonics Toolkit (ATK) plugins**

ATK es un conjunto de complementos para la DAW de Reaper, que incluye diferentes codificadores, transformadores, decodificadores y herramientas necesarias para el proyecto.

- **Camara Cardboard**

Aplicación útil para capturar imágenes 360° mediante el móvil.

- **FB360 Spatial Workstation plugins**

Contiene un conjunto de plugins para conseguir la espacialización y mezcla de audio y video espacial 360°. FB360 se puede emplear usando Pro Tools HD o Reaper. Para este proyecto se va a utilizar con Reaper.

- **AMBEO A-B Format Converter**

Es el complemento específico de los micrófonos de audio espacial Sennheiser que realiza de conversión de Formato A a Formato B, para conseguir tener el audio en la forma estándar.

- **Adobe Photoshop CS6**

Es un reconocido programa de edición de imagen, implementado en el proyecto para conseguir el tamaño de píxeles y el recorte de imágenes necesario para crear las imágenes virtuales 360°.

- **Wondershare Filmora**

Programa de edición de video, utilizado para añadir una duración de tiempo a las imágenes y así conseguir el formato necesario para la reproducción en el complemento VR Video Player.

2.3 DISEÑO DE PAISAJES SONOROS

Se ha facilitado un laboratorio del Edificio el Sario para realizar las diferentes pruebas de paisajes sonoros virtuales.

2.3.1 JUEGO DE COLORES

I. Introducción

En esta primera prueba se pretende evaluar la eficacia de la tecnología Ambisonics. Para ello, se emiten a cada altavoz sonidos grabados a través de un micrófono monofásico y se recogen mediante el micrófono Ambisonic. Además, se quiere observar si el espectador es capaz de discernir por qué lado llega el sonido espacial, una vez convertido a binaural, ayudándose de la imagen visual.

II. Escenario

El primer escenario está compuesto por el equipo necesario para la grabación y reproducción del sonido espacial.

Por un lado, se conecta un micrófono monofásico a la interfaz de audio y este equipo a su vez va conectado a un ordenador. Así, la señal analógica se convierte a digital mediante el interfaz de audio y se recoge en el ordenador, permitiendo la entrada mediante el programa Focusrite y pasando a recogerse la grabación en Reaper.



Figura 4. Instrumentación hardware y funcionamiento del 1º escenario

Por otro lado, desde el ordenador se lanza la reproducción a los 8 altavoces pasando por el interfaz de audio en sentido contrario. Los altavoces formarán una circunferencia con radio de 1,56 m entorno a un eje central donde estará ubicado el micrófono Ambisonics, creando un escenario horizontal de 360° y estarán equiespaciados.

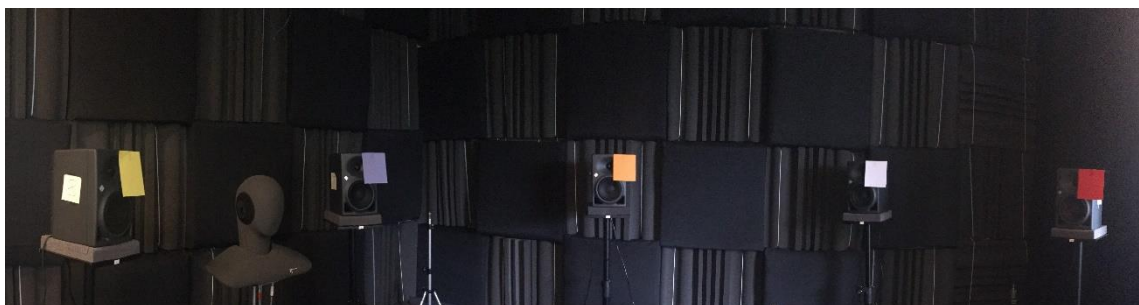


Figura 5. Ubicación altavoces en el laboratorio para representar 1º escenario

El micrófono Ambisnoics estará a una altura de 1,30 m, al igual que los altavoces, e irá conectado al zoom F4, donde se guardarán las grabaciones dentro de la tarjeta SD incorporada.

III. Procedimiento

Para comenzar, se han grabado señales individuales nombrando 8 colores diferentes. Para ello, se ha creado una plantilla en Reaper con ocho pistas en las que se tienen como entrada el micrófono y se recoge en cada pista el sonido de un color diferente.

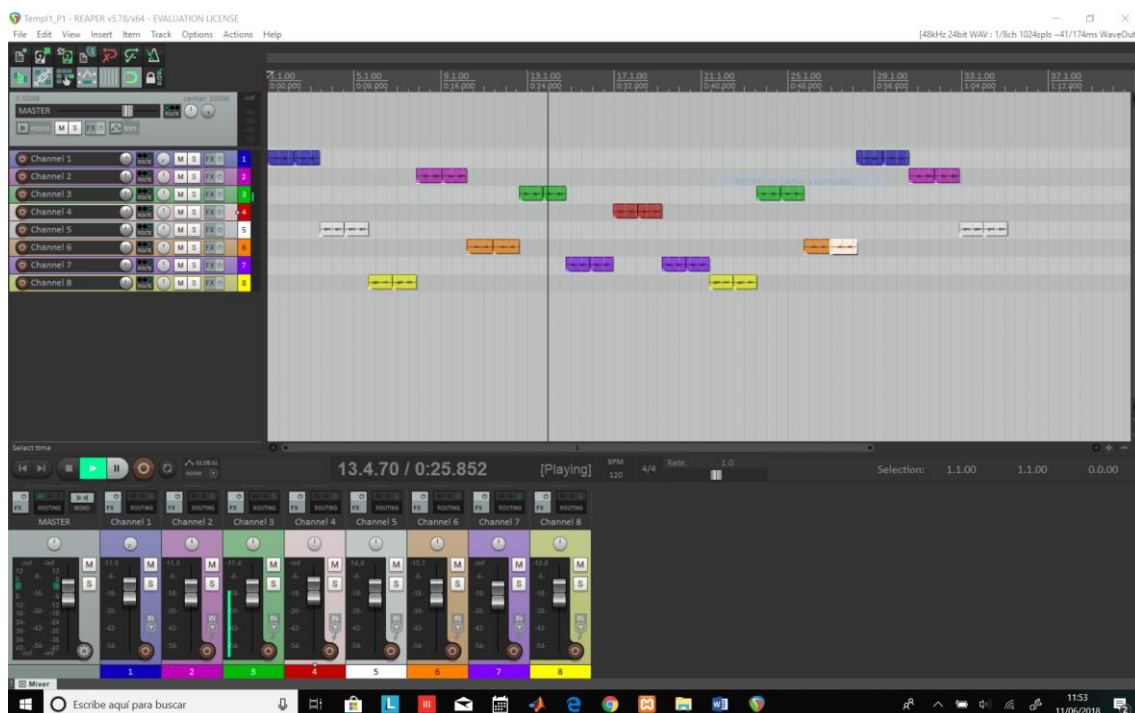


Figura 6. Plantilla Reaper con sonidos de 8 colores

A continuación, se ha configurado para que cada pista se reproduzca por un altavoz diferente. Colocando estos fragmentos grabados en diferente tiempo, hemos creado un audio completo en el que se irán diciendo los colores aleatoriamente. De esta forma, iremos escuchando los colores alternativamente saliendo por los 8 altavoces.

Este audio, se ha recogido mediante el micrófono Ambisonics situado en el centro de la circunferencia con los altavoces. La zoom conectada a este micrófono la hemos configurado en formato FuMa para recoger 4 canales de sonido directamente

en formato-B. Después, hemos pasado esta grabación al ordenador y se ha continuado con la siguiente parte.

En cuanto a la parte visual, se ha creado una imagen ficticia mediante Photoshop con tamaño de 6000x3000 pixeles (relación equirectangular 2:1) para preservar la calidad óptima que se necesita, para posteriormente convertirla a imagen esférica 360°. En ella, se han incorporado ocho circunferencias con cada uno de los colores.



Figura 7. Diseño imagen virtual con 8 circunferencias de colores

Seguidamente se ha introducido esta imagen en una plantilla de Photoshop la cual convierte la imagen rectangular en formato esférico, de forma que al subirla a las plataformas (Facebook, Youtube etc.) se puede ver esférica de 360°.

Para finalizar con la parte relacionada con la imagen, se ha utilizado el programa Wondershare Filmora para darle una cierta duración a la imagen, consiguiendo crear un video necesario para la unión final con el audio espacial.

Volviendo a la parte fundamental del proyecto, el sonido Ambisonico, se ha creado una segunda plantilla en Reaper acompañada de los plugins de FaceBook360.

En este apartado, se mezcla el audio espacial junto con el video de forma que el espectador pueda visualizar las circunferencias junto con los sonidos de los colores provenientes de una esfera de sonido ficticia conforme desliza la pantalla. Esta sensación sonora se escuchará a través de los auriculares, ya que, hemos decidido reproducirlo en audio binaural por ser el método de escucha más común para el usuario.

Como breve resumen, las etapas de flujo de trabajo [8] que se emplean en la plantilla de Reaper para conseguir la idea anterior son:

- **Etapas de Diseño:** En esta etapa se realiza el diseño de sonido y mezcla, utilizando los plugins FB360 Spatialiser y el de Control junto con el reproductor VR para mezclar en binaural en tiempo real.
- **Etapas de Entrega:** Aquí se realiza el proceso para la entrega de contenido. La mezcla de DAW está codificada en un formato que el motor de audio FB360 utiliza para reconstruir una mezcla binaural en tiempo real en el dispositivo de destino. Todos los archivos de audio y metadatos se comprimen en un solo formato de archivo para garantizar el bajo uso de la CPU y la memoria. También se puede entregar la mezcla en otros formatos compatibles con video 360.
- **Reproducción:** Dependiendo del formato de salida, la mezcla puede implementarse en diferentes plataformas de video 360.

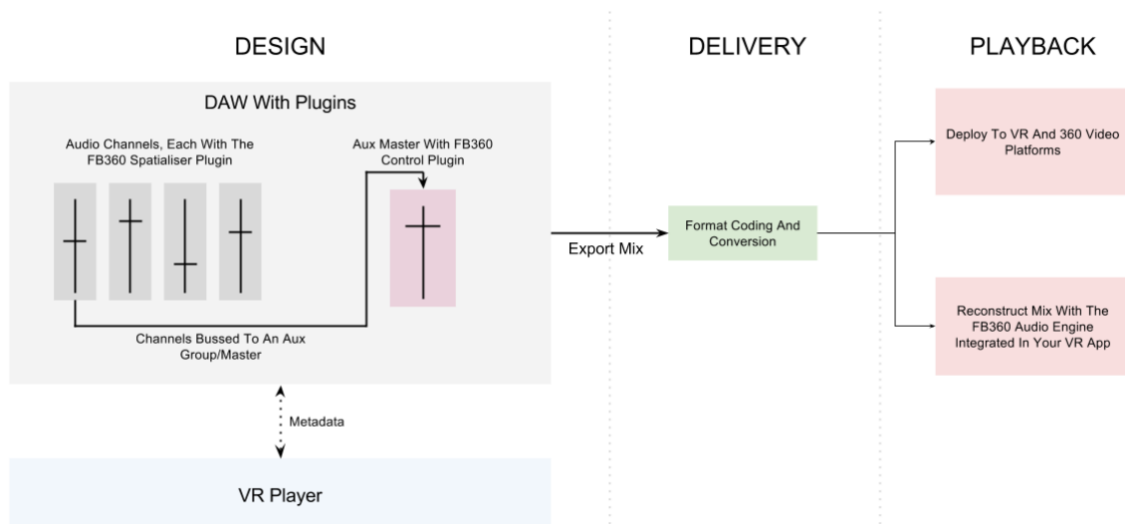


Figura 8. Etapas de flujo de trabajo (FB360 Spatial Workstation guide)[8]

Para comenzar, se ha creado una primera pista introduciendo el audio de 4 canales en formato-B FuMa. En ella, se ha añadido la función FB360 Spatialiser la cual posicionara el video en 3D junto con el audio Ambisonics. Se considera como un panner avanzado que le ayuda a ubicar el audio en el espacio 3D con algoritmos binaurales y de simulación de sala. El complemento panner codifica el audio en ambisonics con metadatos que luego se usan para construir la mezcla completa.

Es importante tener en cuenta que Spatial Workstation [8] trabaja con el formato ambix de 2º orden y ese es el formato en el que se exporta la mezcla de audio final. Por ello, como a la entrada se va a implementar Ambisonic de 1º orden, esto puede hacer que la señal suene más borrosa y difusa. Para contrarrestar esto, se añade un ajuste de ganancia para optimizar la señal en la reproducción, aplicando un aumento de ganancia adicional a los canales XYZ. Si quisiéramos la salida del proyecto en 1º orden, se puede deshacer configurando el control deslizante de envolvente en 0,27.

Un dato a tener en cuenta es que, a la hora de espacializar el video con el audio, ha sido necesario rotar la grabación en formato B 180º en el plano horizontal, para que coincidiera con los colores de la imagen.

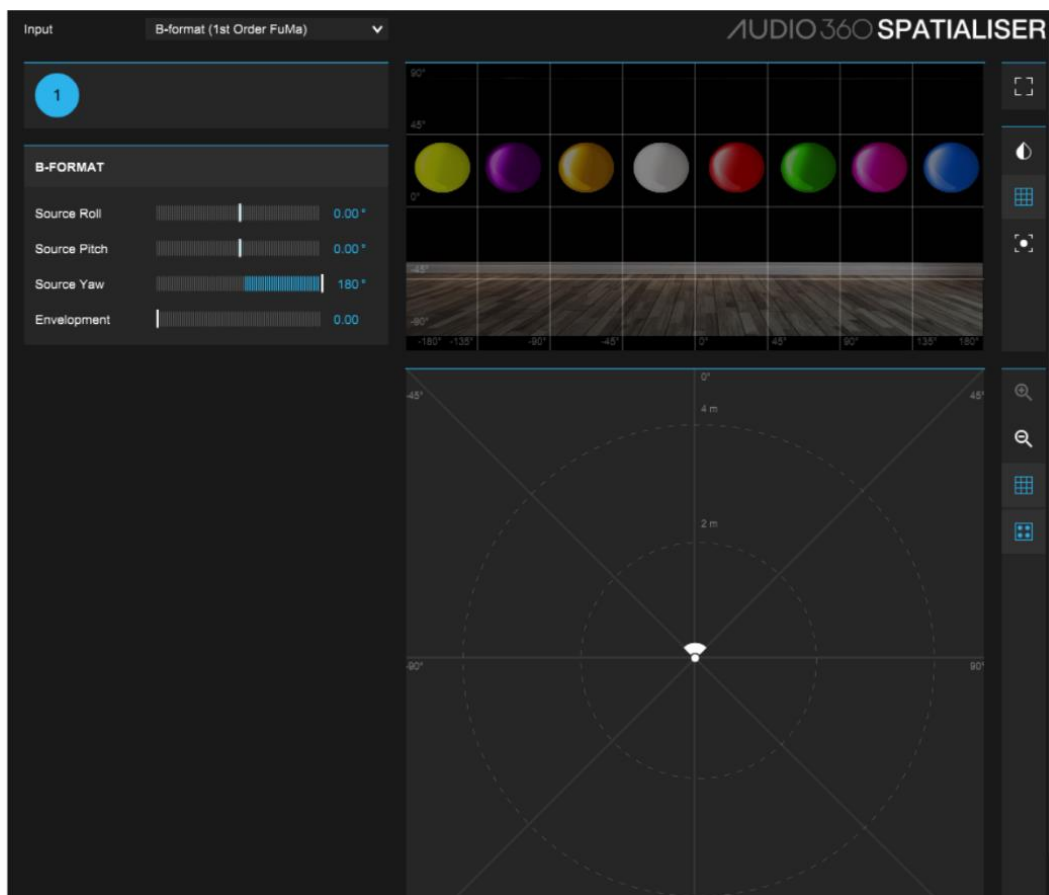


Figura 9. Complemento FB360 Spatialiser Juego colores

Para ver el video en 3D se ha utilizado el plugin VR Video Player. Es una aplicación independiente [8] que se comunica con el DAW para garantizar una buena sincronización con el proyecto. Esto le permite mezclar directamente el video VR mientras se obtiene una vista previa del video en 360°.

Se pueden utilizar dos formas diferentes: Modo Esclavo el cual sincroniza directamente la reproducción de video con su DAW, que ha sido el utilizado en esta prueba. Y por otro lado, Modo Autónomo, que es independiente de su DAW y es una vista previa de un archivo .tbe codificado con un video.

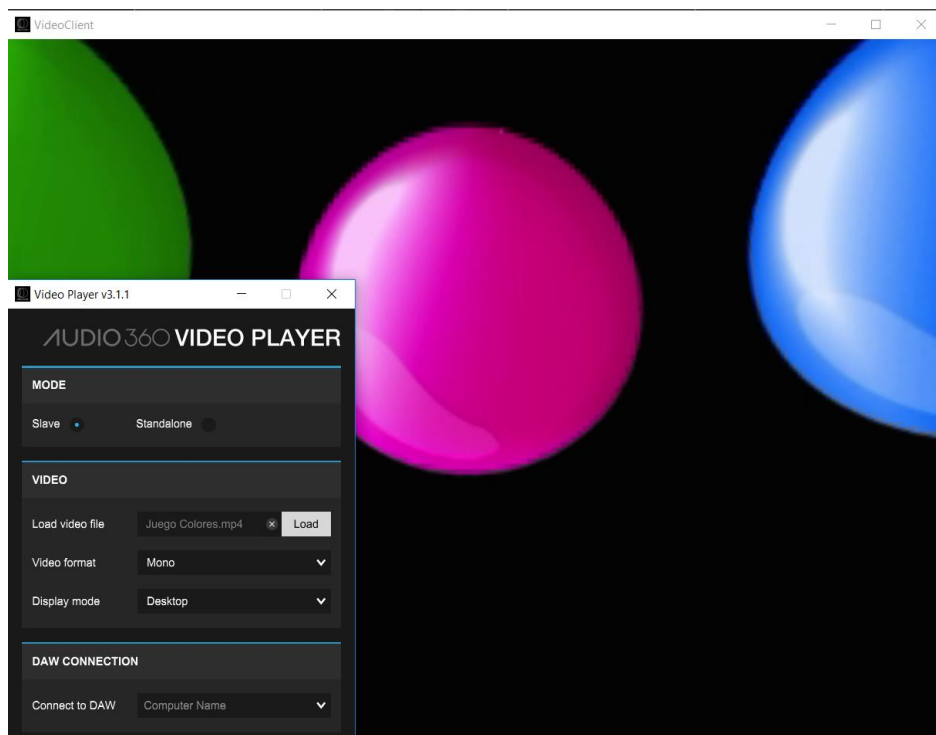


Figura 10. Complemento VR Video Player Juego colores

A continuación, el audio de la primera pista se enruta a “3D MASTER” donde se masteriza la mezcla 3D. Se incluye la función FB360 Converter, que sirve para [8] rotar la mezcla cuando se quiere seguir lo que está sucediendo en la escena o cuando la mezcla está incorrectamente orientada en relación con el video y necesita ser corregido. Este complemento debe ser instanciado entre el plugin Spatialiser y el de Control.

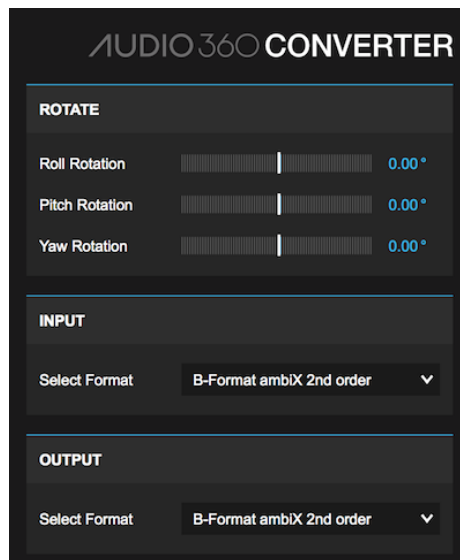


Figura 11. Complemento FB360 Converter

En esta segunda pista, se introduce también el plugin FB360 Mix Loudness. Como es imposible predecir donde mirara el oyente, no se puede determinar en la etapa de mezcla el volumen preciso que escucharía el oyente. Por ello, este complemento sirve para [8] aproximar el volumen general de su mezcla. Se dan unas pautas para la sonoridad máxima posible y el valor máximo real de su mezcla, como si el oyente siempre estuviera mirando en la dirección más alta.

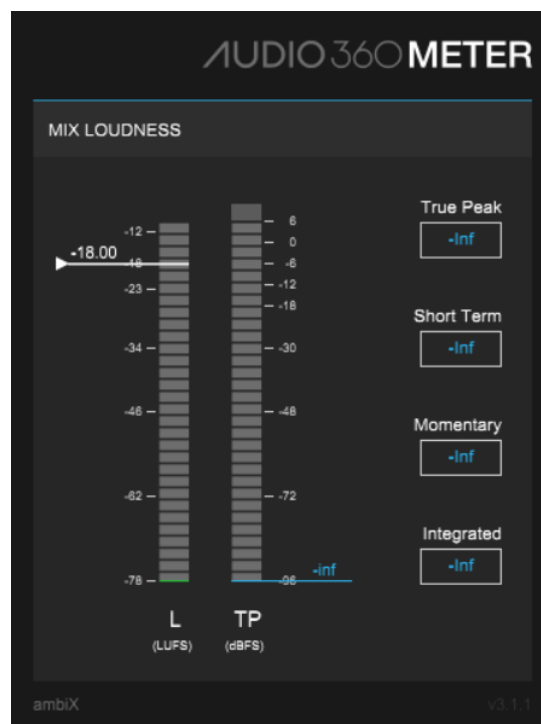


Figura 12. Complemento FB360 Mix Loudness

El tipo de volumen indica el intervalo de tiempo durante el cual se calculó el volumen. Así una ventana más corta significa que el valor del volumen responde rápidamente al audio y una ventana más larga responde lentamente (cerca de un valor de volumen medio). La siguiente tabla muestra el tamaño de la ventana para cada tipo de volumen:

<i>Tipo de sonoridad</i>	<i>Tamaño de ventana</i>
Momentáneo	0.4 segundos
Término corto	3 segundos
Integrado	todo el tiempo

Figura 13. Tabla con el tamaño de ventana según cada tipo de sonoridad [8]

Se ha incorporado una tercera pista llamada “CONTROL PLUGIN” que se encargará de controlar la reproducción y la decodificación binaural. Facilita la comunicación con el reproductor de video VR y controla los parámetros para el modelado de salas. Es decir, controla globalmente el tamaño de la sala, de forma que las reflexiones de modelado de sala se mezclan en la salida Ambisonics del plugin Spatialiser.

Además, se proporcionan los controles de enfoque y orientación del oyente, para mezclar y obtener una vista previa del audio. Esto se aplica en tiempo real durante la reproducción del video.

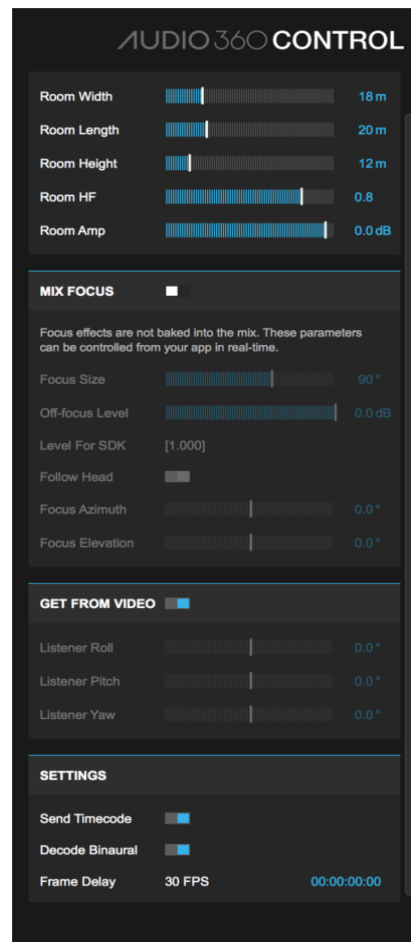


Figura 14. Complemento FB360 Control

Una vez conseguida la mezcla, se ha exportado en un archivo .wav el audio espacial con 8 canales, según las indicaciones dadas en la guía.

El último paso consiste en unir la mezcla de audio junto con el video 360°. Para ello, se necesita un codificador externo que convierte la mezcla de su DAW en una variedad de formatos que pueden subirse a plataformas de video 360°. En este plugin, se ha introducido el audio mezclado y el video y, por último, se ha obtenido el video final 3D en formato .mp4.

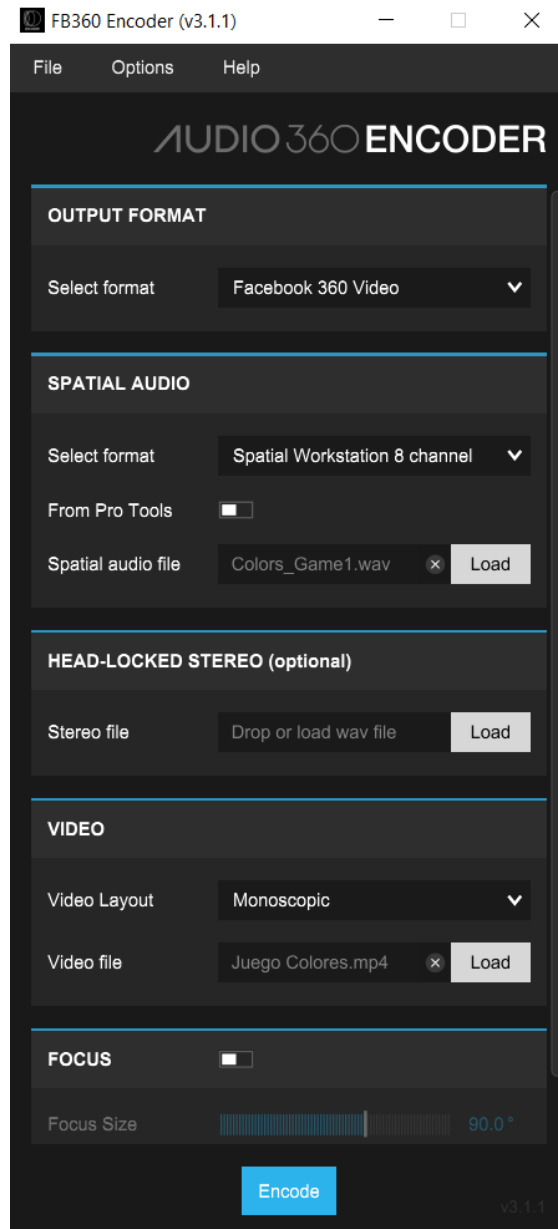


Figura 15. Complemento externo FB360 Encoder

IV. Resultados

Se comprueba como efectivamente, los sonidos reproducidos mediante los 8 altavoces son captados con una gran precisión por el micrófono Ambisonics de Sennheiser Ambeo. Se puede ver como al volver a reproducir los 4 canales guardados en Formato B mediante los altavoces, no se aprecia apenas diferencia para el oído humano entre los dos audios.

Además, observamos como mediante el plugin Spatialiser se posiciona correctamente el video junto con el audio Ambisonics, dando esa sensación de sonido envolvente al espectador.

Por último, se identifica con claridad como en el video 3D codificado, al escuchar el sonido de un color, si se va deslizando la pantalla se tiene sensación de que el sonido rota y llega por el lado izquierdo o derecho, hasta situarse en frente.

2.3.2 JUEGO SELVA

I. Introducción

Este segundo experimento se basa en crear una representación de sonido espacial a través de sonidos monofásicos, recogidos mediante la tecnología Ambisonics. Para la posterior evaluación, de observar si el espectador es capaz de discernir la ubicación con la que llegan los diferentes sonidos de la selva mediante altavoces y convertidos a binaural, pudiendo ayudarse de la imagen de la pantalla.

II. Escenario

En este escenario, se necesita conectar el ordenador a la interfaz de audio, para lanzar los sonidos de los diferentes elementos de la selva desde el programa Reaper hacia los altavoces. Una vez se convierta la señal digital a analógica, pasará del interfaz de audio hacia los 8 altavoces y se reproducirá el sonido.

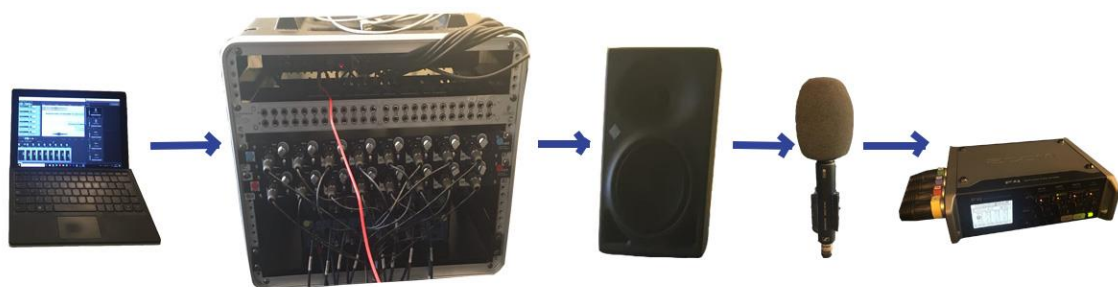


Figura 16. Instrumentación hardware y funcionamiento del 2º escenario

La disposición de los altavoces será la misma que en el primer experimento, es decir, estarán colocados a la misma altura (1,30 m) formando una circunferencia y separados los mismos ángulos, creando un escenario horizontal de 360°.

El micrófono AMBEO irá situado en el centro de la circunferencia a la misma altura que los altavoces. Al micrófono estará conectado el dispositivo zoom donde se captarán las grabaciones en su tarjeta SD.

III. Procedimiento

En primer lugar, se han escogido 8 audios monofásicos correspondientes a sonidos de la jungla. Estos sonidos se han incorporado en Reaper, situando cada audio en una pista diferente. De esta forma, se ha configurado para que cada pista se reproduzca por un altavoz diferente, creando una sensación esférica envolvente.

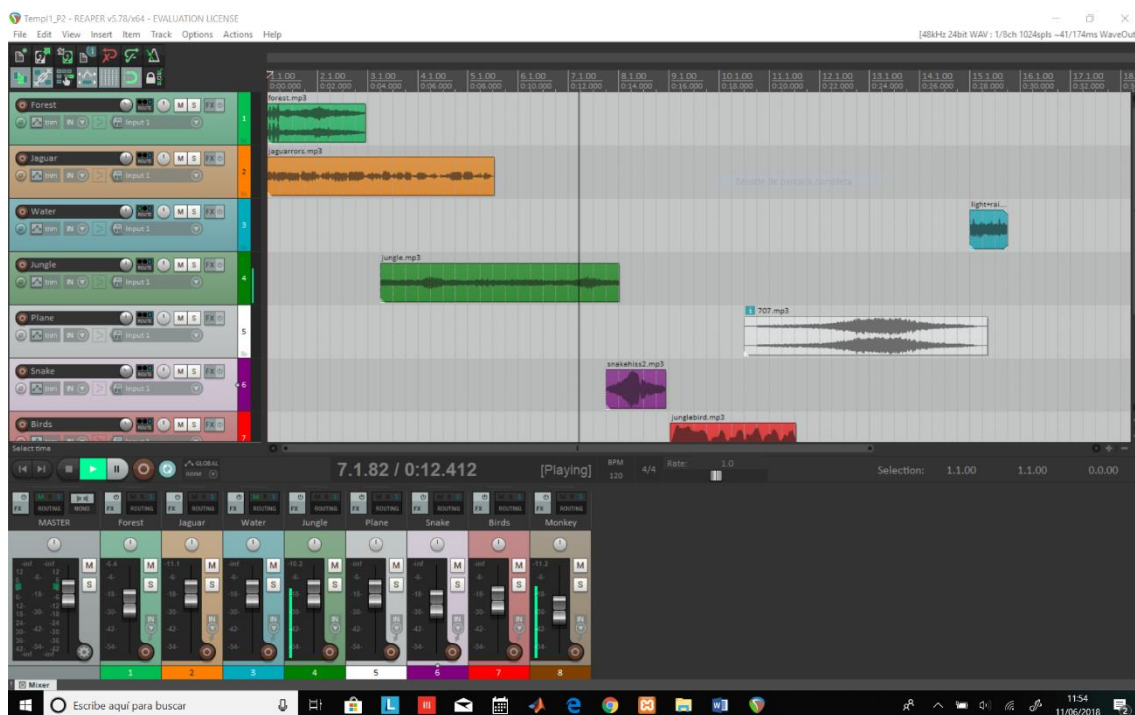


Figura 17. Plantilla Reaper con 8 sonidos de elementos de la selva

El audio completo con los fragmentos de todos los animales y elementos de la jungla se ha grabado mediante el micrófono Ambisonics. La zoom conectada al micrófono se ha configurado para recoger 4 canales tanto en Formato A como en Formato B con convección Ambix.

El siguiente paso, ha sido pasar estas grabaciones al ordenador para crear dos pruebas diferentes; reproducción binaural y mediante altavoces del sonido espacial.

En cuanto a la parte visual, se ha creado una imagen con los diferentes elementos de sonido (jaguar, mono, serpiente, avión etc.) ubicándolos en donde irían situados en el espacio real, en este caso, en la ubicación donde estaba cada altavoz por el cual ha salido el sonido. Para ello, se ha empleado el programa de edición de imagen Photoshop.

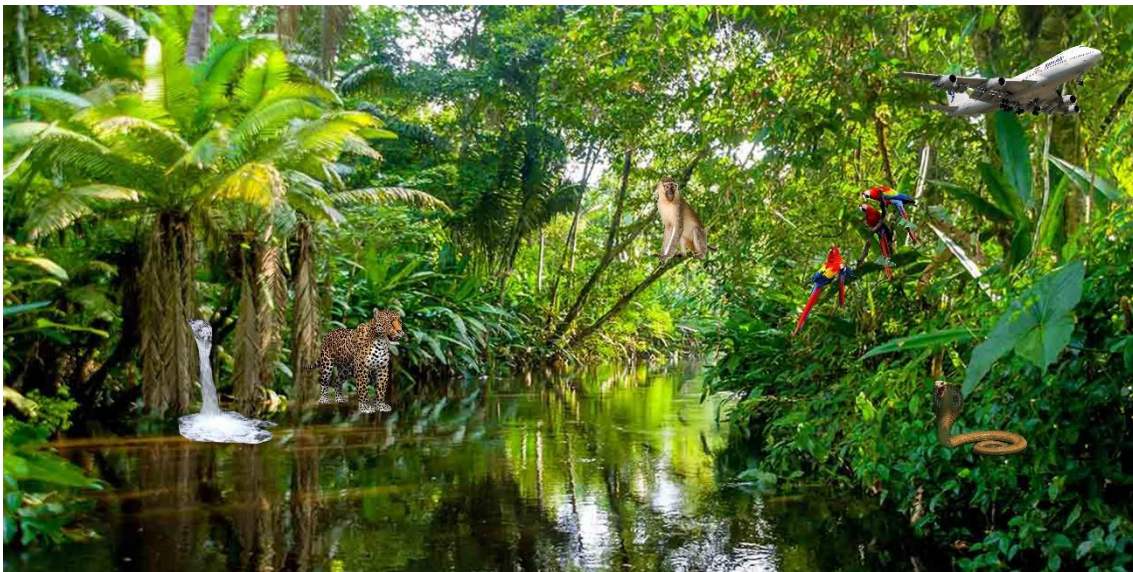


Figura 18. Diseño imagen virtual jungla con 8 elementos de sonido

Después, se ha incluido en la plantilla de 6000x3000 pixeles para ponerlo en formato 360° y finalmente se le ha añadido una duración para crear un video.

Para comenzar con la espacialización, se ha creado una segunda plantilla, en la cual se ha añadido el audio en Formato-B de Ambix. En la primera pista, se ha introducido el plugin FB360 Spatialiser, el cual posicionara el video en 3D junto al audio Ambisonics. Recordar, que para trabajar con ello el programa lo convertirá el audio a Ambix de 2º orden.

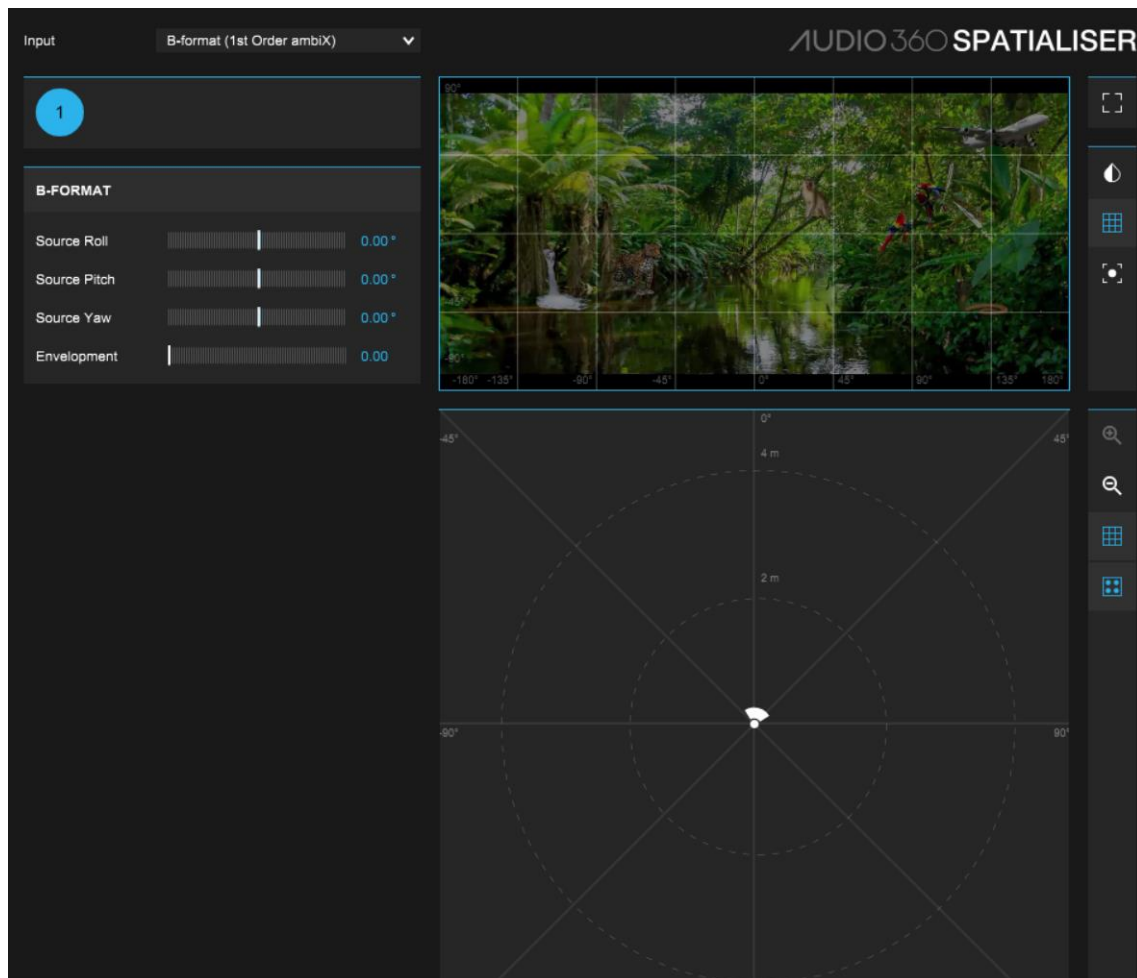


Figura 19. Complemento FB360 Spatialiser Juego selva

A continuación, se observa el video en 3D mediante el plugin VR Virtual player. Se crea una segunda pista “3D MASTER”, donde se masteriza la mezcla 3D de la primera pista. Además, se añade una tercera pista que se encarga de controlar la reproducción binaural.

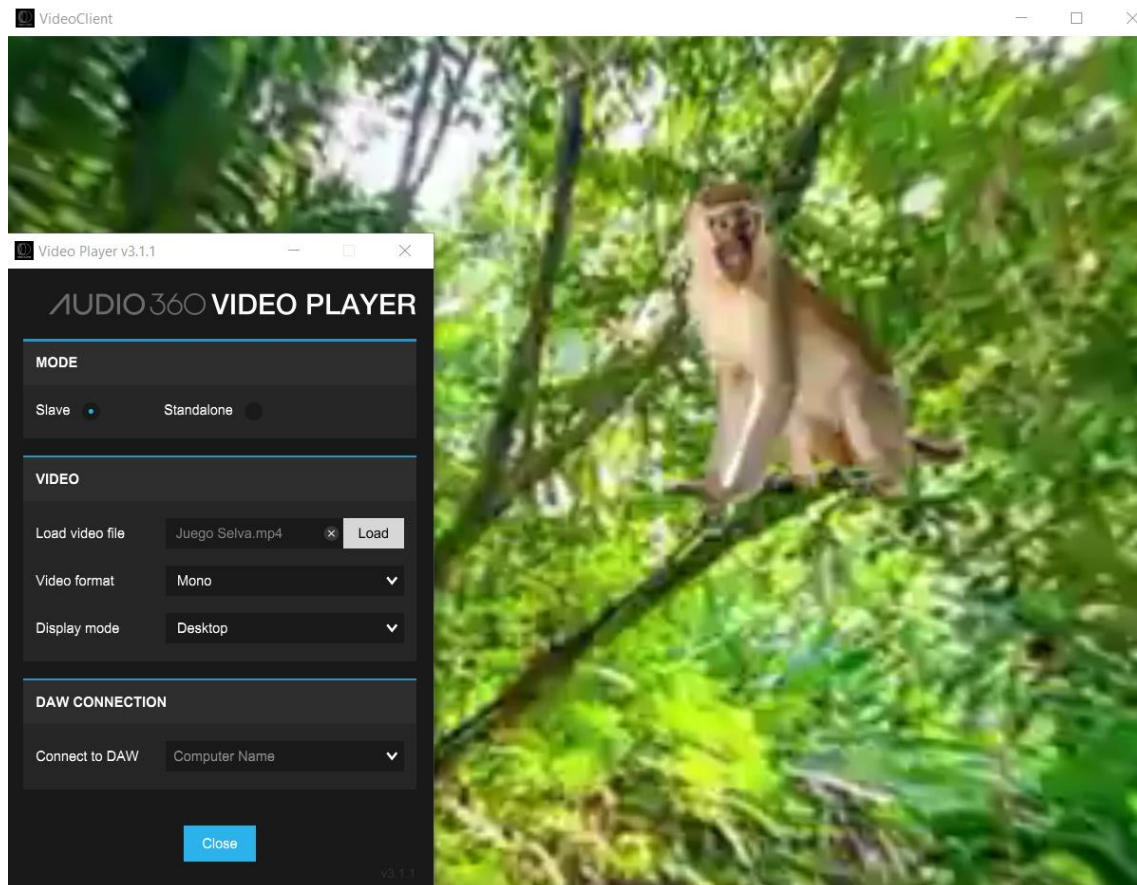


Figura 20. Complemento VR Video Player Juego selva

Por último, se ha realizado el mismo proceso explicado en el primer escenario para conseguir el video en 3D con la reproducción binaural. Así, concluimos con la primera prueba, donde el espectador, puede escuchar mediante auriculares el sonido envolvente 3D y visualizar el video de la selva para la inmersión completa.

Para la segunda prueba, se ha creado una nueva plantilla de Reaper en la cual se introduce el audio de 4 canales grabado con el micrófono AMBEO en formato A. Como se ha explicado anteriormente en la teoría, el formato estándar de Ambisonics con el que se trabaja es el formato B. Por ello, ha sido necesario añadir un plugin que realice la conversión entre Formato A y B. En el proyecto hemos utilizado el complemento software propio del micrófono AMBEO.

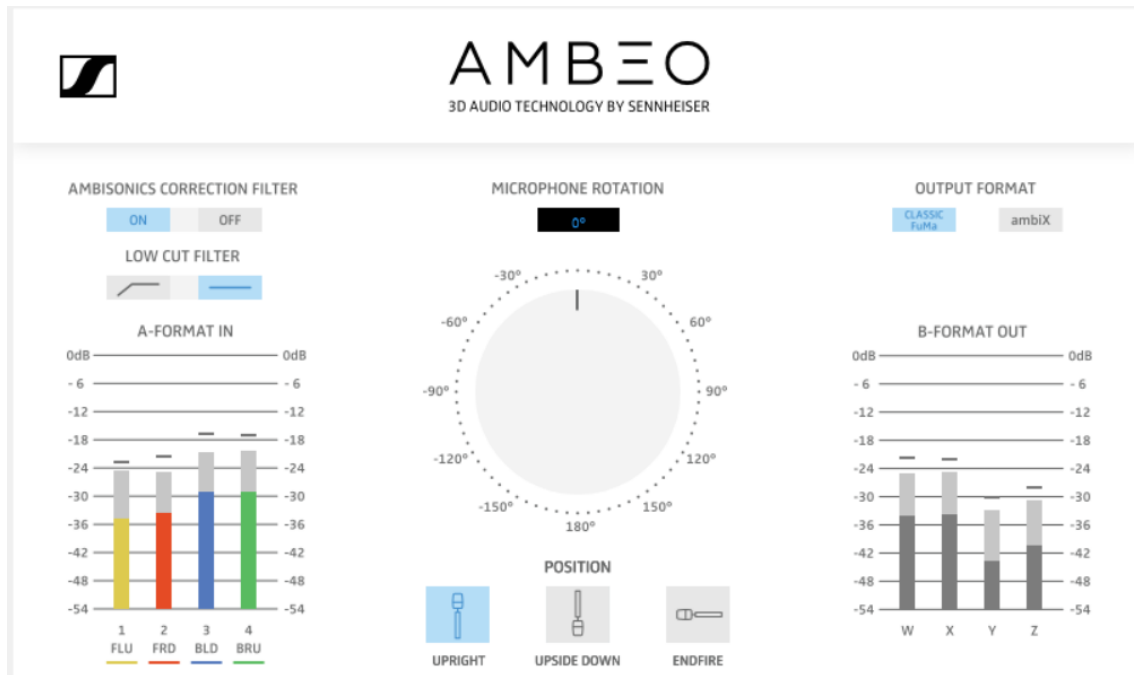


Figura 21. Complemento AMBEO A-B Format Converter

Una vez se tiene el audio de 4 canales en formato B, en este caso hemos escogido la convección FuMa, se necesita reproducirlo mediante 8 altavoces. Sabiendo como se ha visto en la teoría, que la disposición de los altavoces está ubicada creando un escenario horizontal de 360° (sistema pantofónico), se necesita un decodificador pantofónico.

Para ello, se ha empleado el [7] decodificador pantofónico de 1º orden de los plugins de Ambisonics Toolkit (ATK). Así, conseguimos que el espectador pueda escuchar un sonido esférico 3D, ayudándose del video, para la recreación e inmersión completa de la selva.

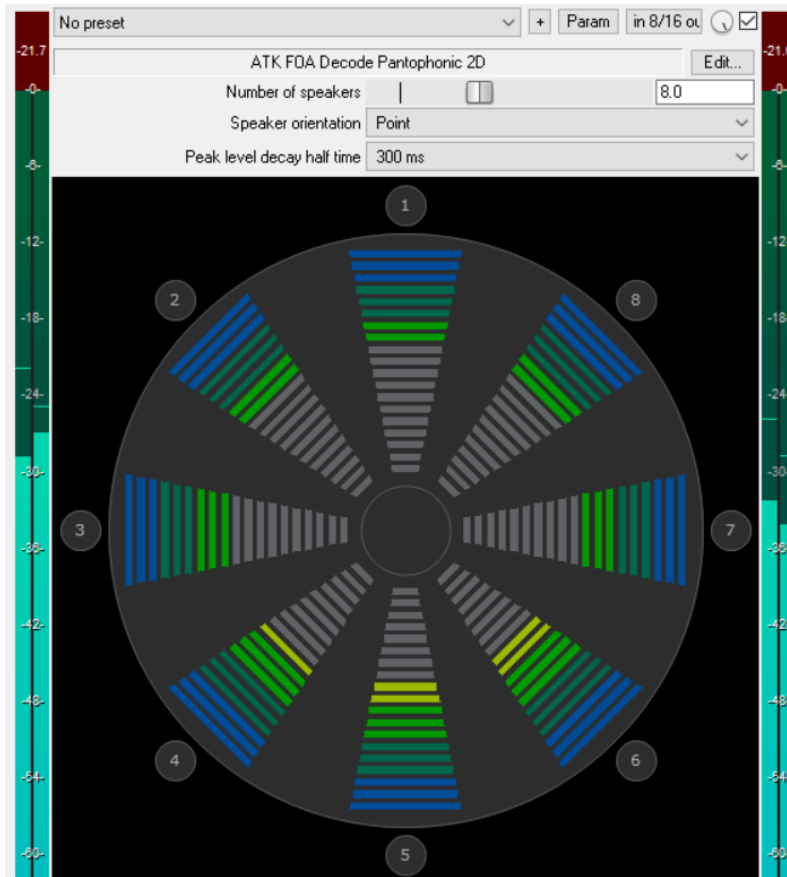


Figura 22. ATK plugin FOA Decode Pantophonic 2D [7]

IV. Resultados

Como resultados de esta segunda prueba, se comprueba como en ambas representaciones, tanto en binaural como con altavoces, se tiene sensación de audio espacial. Sin embargo, cabe decir, que resulta más difícil diferenciar con exactitud de que dirección llega cada sonido de la selva, ya que, se ha reproducido en algunos momentos más de una fuente a la vez.

A modo de apreciación, se ha podido ver como cuando se tienen dos fuentes sonoras muy distanciadas entre sí (enfrente un sonido del otro p.e) se disciernen con mayor facilidad que cuando están sonando dos fuentes cercanas entre sí.

Por último, cabe destacar que el plugin conversor A-B de AMBEO junto con el complemento decodificador de AmbisonicsToolkit consiguen una muy buena reproducción de la escena de la selva mediante altavoces, una vez introducido el audio Ambisonics.

2.4 GRABACIÓN Y REPRODUCCIÓN DE PAISAJES SONOROS

2.4.1 APLICACIÓN CAMINO FLUVIAL

I. Introducción

Se ha decidido como ubicación para grabar paisajes sonoros reales, captar los puntos más interesantes del Camino Fluvial de Pamplona. Posteriormente, para dar una utilidad a dichos videos, se ha creado una aplicación con los diferentes tramos del Camino Fluvial, en los que, clicando en los puntos significativos de cada tramo, se pueden ver los videos 3D con audio espacial incorporado.

II. Escenario

En esta ocasión el escenario será dinámico, por ello, se intenta llevar únicamente el equipo imprescindible. Se ha necesitado el micrófono Ambisonics, el trípode para sujetarlo a la altura correcta y el equipo zoom F4 conectado al micrófono para guardar la información. En ocasiones en las que ha habido bastante viento, se le ha añadido una funda corta vientos al micrófono.



Figura 23. Instrumentación hardware del tercer escenario

Se ha realizado el recorrido a las orillas del río desde Barañain hasta Burlada y se han capturado las zonas simbólicas; visualmente atractivas con sonidos

interesantes (principalmente puentes). En total se han escogido 10 puntos importantes de entre todas las grabaciones realizadas.

III. Procedimiento

En cada punto de captura, se han hecho generalmente dos grabaciones, la primera configurando la zoom para recoger sonido en Formato A y la segunda en Formato B Ambix. Las duraciones aproximadas de cada audio se han recogido de 1 minuto.

Acto seguido, se ha capturado la imagen 360° en el mismo punto donde se situaba el micrófono con la aplicación Cámara Cardboard.

Una vez obtenidos los audios e imágenes necesarias se han volcado al ordenador. El siguiente paso ha sido modificar la imagen para crear el video 3D necesario para añadirle el audio espacial. En este caso, a diferencia de los anteriores, como la imagen había sido capturada con tamaño 9964x1647 se ha conformado de forma panorámica 360°, ya que modificándola a forma esférica 360° se veía deformada.

Por lo que al audio espacial respecta, se ha creado una plantilla Reaper por cada una de las 10 ubicaciones. En ella, se ha incorporado una pista para el primer audio recogido en Formato A y otra para el Formato B ambix. Se han analizado ambas pistas de sonido escuchando los sonidos característicos y seguidamente, se han editado, recortando todos los momentos de viento y sonidos menos interesantes, para dejar únicamente los 30 segundos más representativos uniendo ambas pistas.

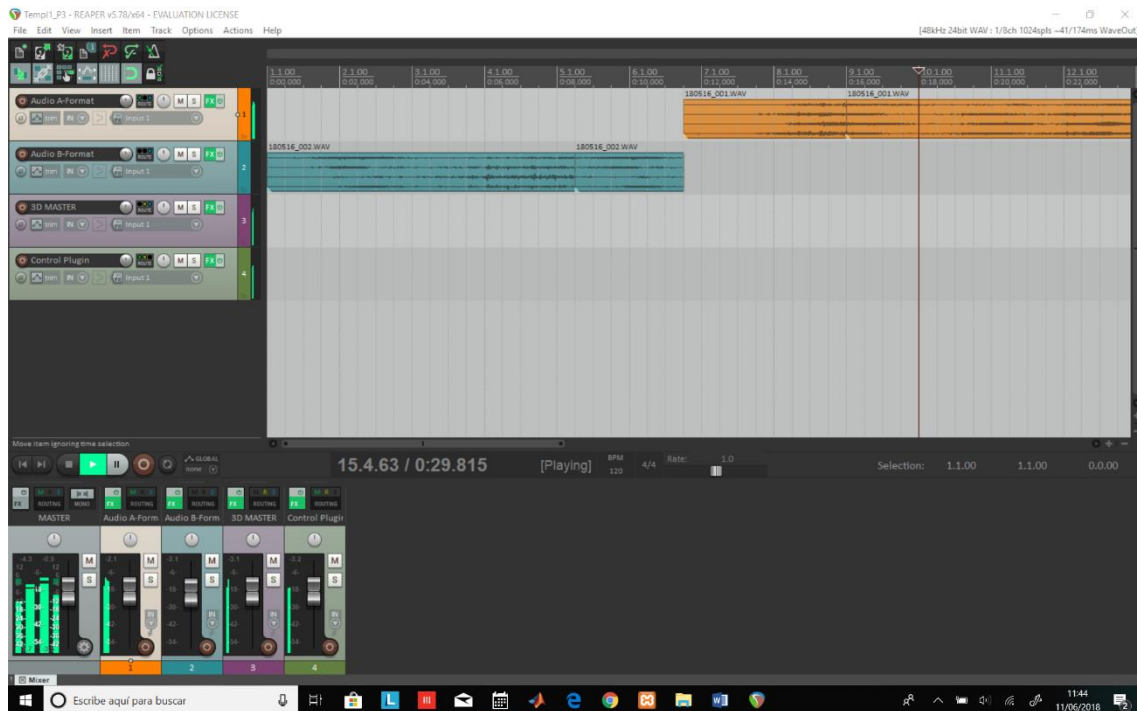


Figura 24. Plantilla Reaper Camino Fluvial Ubicación 1

A continuación, se ha comenzado con la especialización a ambas pistas, pero antes de ese paso, se ha incorporado en la primera pista el convertor AMBEO para modificar el audio de Formato A formato B ambix y así tener ambas pistas con la misma configuración. Con el plugin FB360 Spatialiser, se posicionará el video 3D junto al audio Ambisonics como hemos comentado en las anteriores pruebas.

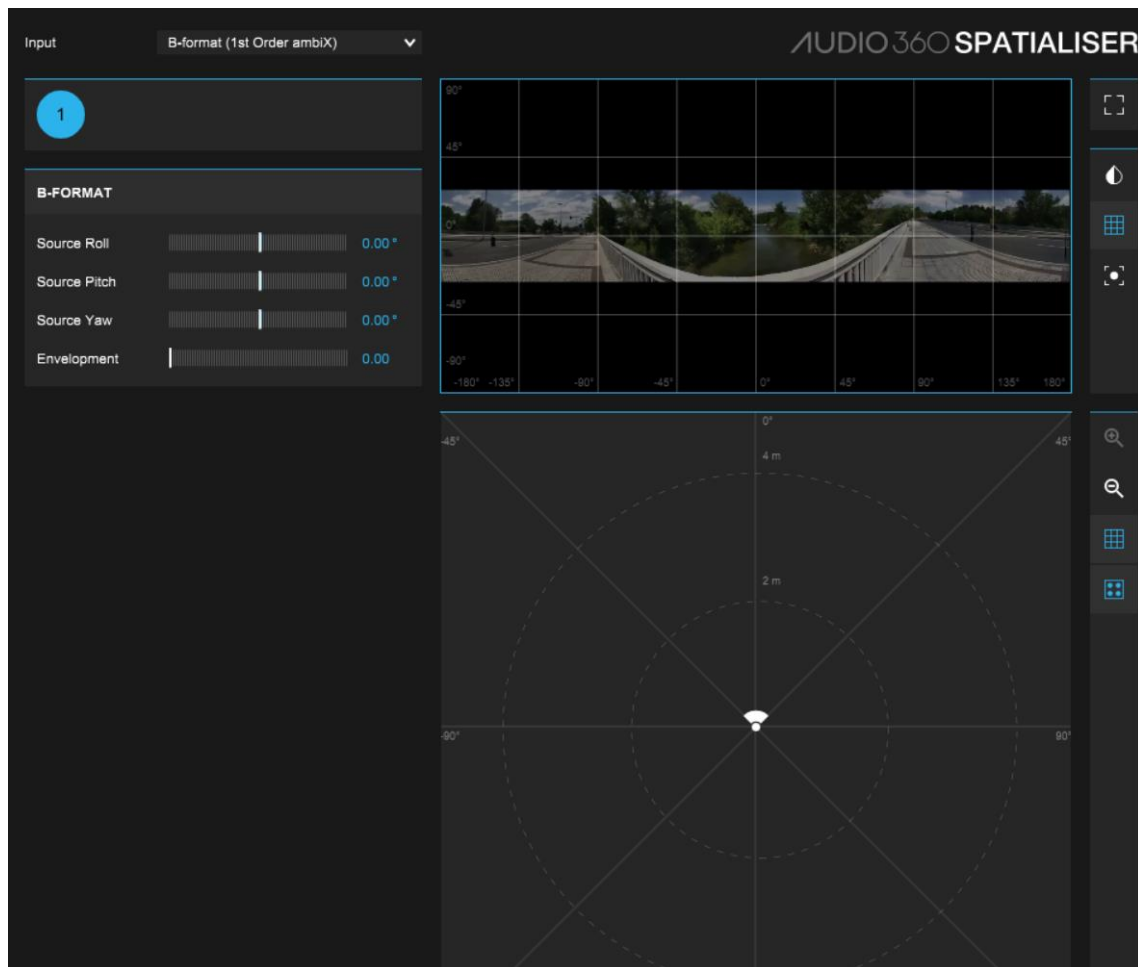


Figura 25. Complemento FB360 Spatialiser con el Camino Fluvial

Justo después, se ha visualizado el video en 3D mediante el plugin VR Virtual Player y se ha creado la siguiente pista “3D MASTER”, donde como ya sabemos, se masteriza la mezcla 3D de las dos primeras pistas. Además, hemos añadido una última pista encargada de controlar la reproducción binaural.

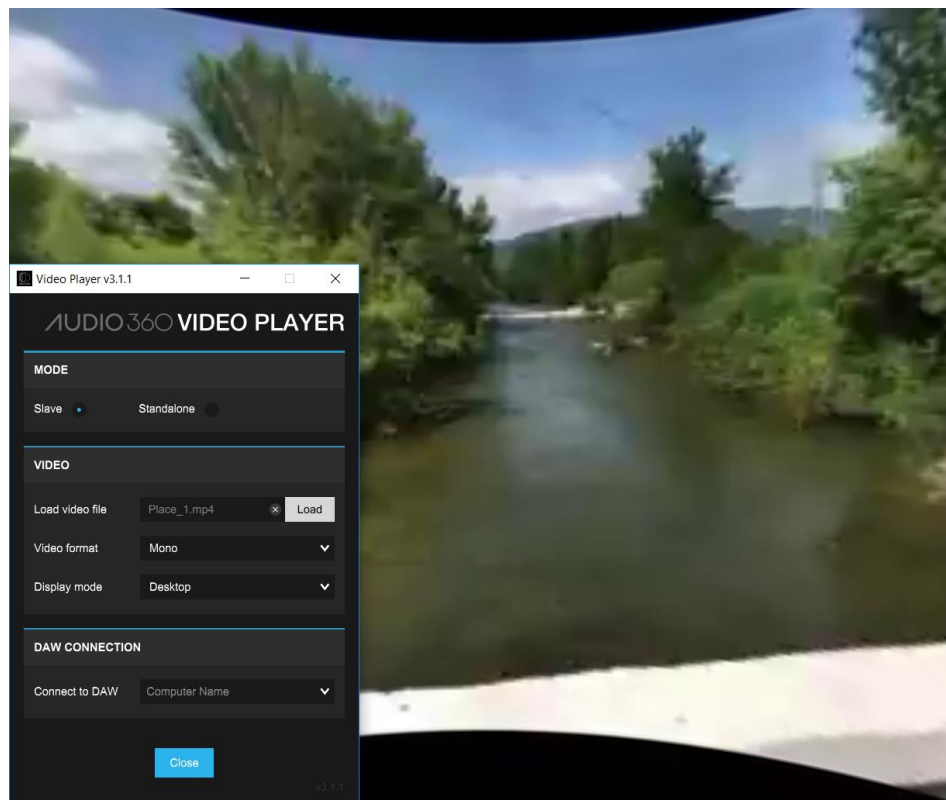


Figura 26. Complemento VR Virtual Player Camino Fluvial

Para concluir con esta parte, se ha unido el audio y video espacial mediante el plugin externo FB360 Encoder, el cual se detalla el proceso en la primera prueba. De esta forma, hemos conseguido 10 videos 3D con reproducción binaural, uno por cada ubicación del Camino Fluvial recogida. Estos videos se han subido a la plataforma Youtube, para que, los usuarios puedan ver en línea los videos 3D con los auriculares mediante una aplicación que hemos creado.

Para comenzar con la aplicación, se ha montado el servidor local XAMPP para la creación de la página web, en donde se ha utilizado el servidor web Apache y los paquetes necesarios para el empleo de los scripts PHP.

Acto seguido, hemos creado una carpeta de nombre 'images' donde se han introducido cuatro imágenes correspondientes a los tramos grabados del Camino Fluvial: Tramo San Jorge, Rochapea, Aranzadi y Magdalena.

En cuanto a la programación web, se ha incorporado un archivo HTML 'index.html' [9] para la introducción en la web de las imágenes y para la buena apariencia y clasificación de los bloques.

```
// Insertar imágenes
```

```




```

Como se puede ver, se han creado cuatro identificadores para diferenciar las distintas áreas de [10] imágenes de manera dinámica, las cuales serán necesarias en Javascript.

A continuación, nos hemos ayudado de un Framework Bootstrap [11] para la utilización de un modal. De esta forma se facilita y hace más visual el desarrollo del código. Se quiere que cuando el usuario clique en un punto del Camino Fluvial, le pueda salir una ventana con el video 3D, por ello, para la creación de la ventana nos ayudaremos del modal.

```
//Modal
```

```
<div class="modal fade" id="myModal" tabindex="-1" role="dialog"
aria-labelledby="ModalCenterTitle" aria-hidden="true">
  <div class="modal-dialog modal-lg modal-dialog-centered" role="document" >
    <div class="modal-content">
      <div class="modal-header">
        <h5 class="modal-title" id="ModalCenterTitle">Camino Fluvial</h5>
        <button type="button" class="close" data-dismiss="modal" aria-label="Close">
          <span aria-hidden="true">&times;</span>
        </button>
      </div>
      <div class="modal-body">
        <iframe width="100%" height="315" id="video" src="" frameborder="0"
allow="autoplay; encrypted-media" allowfullscreen align="center"></iframe>
      </div>
      <div class="modal-footer">
        <button type="button" class="btn btn-secondary" data-dismiss="modal"
>Close</button>
      </div>
    </div>
  </div>
</div>
```



Figura 27. Representación del modal en la aplicación para visualizar el video

Se ha requerido el empleo de un segundo modal modificado, ya que, en una de las ubicaciones, en concreto en la Presa de Harinera, se recogieron dos grabaciones interesantes. Por ello, se han añadido dos rutas y dos iframes diferentes en la parte de HTML para visualizar ambos videos 3D correctamente.

Para el empleo de Javascript, se ha añadido la librería JQuery para facilitar el desarrollo del código. En esta parte, hemos diferenciado 4 apartados, uno por cada imagen, introduciendo los identificadores explicados anteriormente (#img1 p.e). En cada sección, se ha determinado mediante las coordenadas de los pixeles de cada imagen, cuáles son las áreas importantes, en las que cuando el usuario clique la imagen le aparezca la ventana con la reproducción del vídeo. En ese espacio, se introduce una ruta que conduce a Youtube, ya que se han decidido subir los videos a esta plataforma debido a que soporta la reproducción de video y audio espacial. De esta forma el usuario podrá verlo en línea correctamente.

```
// Determinar coordenadas y añadir ruta a video en el modal

<script> $(document).ready(function() {
    $("#img1").on("click", function(event) {
        var x = event.pageX - this.offsetLeft;
        var y = event.pageY - this.offsetTop;
        // alert("X Coordinate: " + x + " Y Coordinate: " + y);

        if ((x<770) && (x>702) && (y<617) && (y>567)){
            ruta="https://www.youtubenocookie.com/embed/DEGWq97MFEw?rel=0?
vq=hd270";
            document.getElementById('video').src= ruta;
            $("#myModal").modal();
        }
    });
}); </script>
```

La ruta ha sido cambiada dinámicamente del modal, para poder reducir el código a lo imprescindible, sin introducir redundancias.

Por último, para que los usuarios y las personas interesadas puedan ver la aplicación se ha buscado un hosting gratuito y se han subido los ficheros desarrollados. De esta forma, todo usuario interesado en el ámbito espacial y en la naturaleza podrá acceder a esta web desde cualquier parte del mundo y no solo desde el servidor local.

2.5 EVALUCACIÓN MEDIANTE TEST SUBJETIVOS DE PAISAJES SONOROS

I. Introducción

Como prueba final, se quería observar la percepción de las personas al sonido espacial Ambisonics y cómo este sonido cambia al convertirlo en audio binaural, en definitiva, valorar los diferentes métodos de reproducción. Para ello, se han realizado diferentes test subjetivos a un grupo de diez personas a cerca de los paisajes sonoros ficticios y reales de los apartados anteriores.

II. Procedimiento

En primer lugar, se ha creado un formulario de Google con 10 preguntas a contestar, para ver las sensaciones que ha habido una vez finalizadas las pruebas y así poder analizar los resultados.

La primera prueba ha sido el juego de colores. Se ha ubicado a la persona en el centro de la circunferencia de altavoces, se le han colocado las gafas 3D y se ha reproducido el audio de los colores mediante los 8 altavoces. Con ello, se quería valorar la sensación espacial inmersiva que tenían.



Figura 28. 1ª prueba en el laboratorio mediante altavoces y gafas 3D

Acto seguido, se les ha enseñado el video del juego de colores mediante la pantalla y los auriculares. De esta forma, se quería ver las diferencias que notaban entre la reproducción Ambisonics y la reproducción binaural. Además, de la sensación de campo envolvente que les producía verlo con las gafas 3D o desde la pantalla.

Para finalizar esta primera parte, se han modificado algunos colores de salida para que no correspondan con el altavoz del cual debería de salir ese sonido del color. Es decir, han sido mezclados los colores y las ubicaciones por las que debería de llegar el sonido de este color. Con esto, se quería ver la capacidad de percepción de las personas, si se guían más por la parte visual o por la sensación de ubicación del sonido.

Para la segunda prueba, hemos empleado el Juego de la selva. Aquí, se ha reproducido el audio mediante los 8 altavoces y se ha visualizado la imagen con las gafas 3D. En este apartado se quiere observar, si al emitir sonido por más de un altavoz mezclando los diferentes sonidos de la jungla, se sigue discerniendo correctamente el sonido emitido por cada fuente sonora.

Para concluir, se han hecho una serie de preguntas a cerca de los videos grabados del Camino Fluvial. Por un lado, se quería ver si se pierde calidad en la codificación final al convertir en audio a binaural y unir el video 3D. Por ello, se ha preguntado a los participantes, si escuchaban mejor el audio espacial directamente desde la plantilla de Reaper o desde el Video final 3D.

Por otro lado, se quería ver si efectivamente los participantes diferenciaban bien por qué direcciones les llegaba el sonido espacial. Para ello, se seleccionaron cuatro de los 10 videos, que tuvieran sonidos característicos y se les hicieron preguntas acerca de por qué dirección creían que venía un sonido específico. Algunas de las preguntas que se hicieron fueron las siguientes:

- ¿En qué dirección va el sonido del avión en el Puente de Miluce?
- ¿De qué lado viene el sonido de la motocicleta en el puente de la Rochapea?
- ¿En qué lado se encuentra la niña lanzando piedras en las pasarelas?



Figura 29. Participante escuchando binauralmente y con gafas 3D el sonido del Camino Fluvial

III. Resultados

Una vez procesadas las respuestas, se ha llegado a los siguientes resultados.

Para comenzar, la totalidad de los participantes han tenido la sensación de sonido inmersivo en la prueba. Esto quiere decir, que la reproducción de audio mediante la tecnología Ambisonics es una muy buena representación.

A la hora de preguntar con cuál de los métodos escuchaban mejor el sonido espacial y tenían mayor sensación envolvente, la gran mayoría, ha escogido la reproducción mediante altavoces con la incorporación de las gafas 3D.

Posteriormente, en el momento en el que un sonido de un color no concordaba con la imagen de donde estaba ubicado, la gran mayoría de los participantes lo detectaba. A pesar de que, cuando el sonido salía de forma contraria o muy distanciada lo diferenciaban muy bien, pero si la modificación se había hecho para que saliera de un altavoz cercano, entonces se prestaba a mayor confusión para diferenciarlo. Se puede decir que, por tanto, los participantes se guiaron más por la sensación de la dirección del sonido espacial, que de la percepción visual.

Pasando al juego de la selva, hemos comprobado como en los instantes en los que no había una única fuente sonora reproduciéndose, a los participantes les costaba más averiguar con seguridad porque lado llegaban los sonidos de la selva. A diferencia del juego de colores en el que únicamente se escuchaba una sola fuente emisora en cada momento y lo diferenciaban muy bien. En esta parte con diferentes fuentes de emisión al mismo tiempo les costaba un poco más.

En cuanto a la última prueba, se han mostrado los videos del camino fluvial seleccionados directamente con el audio espacial desde la plantilla de Reaper y una vez codificado a binaural. La totalidad de los participantes han apuntado que el sonido pierde una notable calidad al ser codificado para obtener el video 3D. Sin embargo, en cuanto a la calidad de la imagen, dicen que aparece apenas inalterada.

En las preguntas relacionadas con la posición concreta donde provenía un sonido específico, se observa como en los sonidos que transcurren un considerable periodo de tiempo (avión p.e) o un sonido puntual pero fuerte (claxon p.e), el oído

los capta y diferencia mejor que sonidos más lejanos con no mucha duración (motocicleta p.e). Además, si el sonido ambiente no es despreciable respecto al sonido puntual, no se diferencia con certeza la ubicación con la que llega al oyente.

3. CONCLUSIONES

- Se han analizado diferentes programas para la creación de videos 360° introduciendo audio espacial Ambisonics, y finalmente, se ha escogido el programa FB360 spatial Workstation with Reaper y los complementos de Ambisonics Toolkit.

El primer software se ha escogido por la compatibilidad con el programa de sonido Reaper, por ser gratuito y por incorporar numerosas funciones para describir, reproducir y codificar un campo de sonido 3D completo. En segundo lugar, nos hemos decantado por Ambisonics Toolkit, ya que, ofrece una variedad de herramientas para incorporar a Reaper como; codificadores, transformadores, decodificadores etc.

- En la evaluación de los complementos encargados de la espacialización se ha visto como se posiciona correctamente el video junto con el audio Ambisonics, dando una sensación de sonido envolvente al espectador.
- A la hora de reproducir el audio binaural, por lo que a la codificación de FB360 respecta, se ha apreciado como una vez realizadas las tres pruebas, ha sido útil y ha servido de forma correcta para las dos primeras escenas sonoras ficticias grabadas dentro del laboratorio. Sin embargo, para el caso del camino fluvial, se ha percibido como en la codificación perdía bastante calidad y nitidez mezclando los sonidos característicos y el sonido ambiente de fondo.
- En la reproducción mediante altavoces, se ha utilizado el decodificador de ATK. En dicha prueba, hemos podido observar como cuando se tienen dos fuentes sonoras muy distanciadas entre sí (enfrente un sonido del otro p.e) se disciernen con mayor facilidad los sonidos, que cuando están sonando dos fuentes cercanas entre sí.
- Se ha creado una aplicación para el uso y provecho de los videos 3D obtenidos a cerca del recorrido del Camino Fluvial.
- Se han conseguido las expectativas de obtener sonido inmersivo, viendo como los participantes diferenciaban correctamente la dirección espacial por la que provenía el sonido.

- Evaluando las preferencias de los participantes, han optado por el uso de altavoces y gafas 3D como mejor sistema de sonido envolvente. Gracias a la tecnología Ambisonics se permite reproducir el sonido espacial mediante altavoces de forma muy precisa. Esto hace que sea muy útil e interesante pensar en esta tecnología para incluirla en futuras aplicaciones. De todas formas, hemos visto durante las pruebas como esta tecnología es válida tanto para reproducciones con altavoces como para reproducciones binaurales.

4. BIBLIOGRAFÍA

- [1] Mario Ledo Hernansanz “Tecnología de registro y reproducción sonora para entornos de realidad virtual” Proyecto Fin de Carrera,UPM, Jun. 2013
- [2] Malham, David G., Myatt, Anthony “3-D sound spatialization using ambisonic techniques “ Computer Music Journal, 19 (4), pp. 58-70. Cited 107 times, 1995
- [3] Jhosimar Aguacía Fisco, Diego Pardo Sabogal “Reconstrucción de Átmósferas Sonoras Tridimensionales” Fac. Ing. De Sonido, Bogotá. 2, Sept. 2013
- [4] Sennheiser AMBEO VR Mic [En línea] Disponible en: <https://en-us.sennheiser.com/microphone-3d-audio-ambeco-vr-mic>
- [5] D.G.Malham “Spatial hearing mechanisms and sound reproduction” Music technology group, University of York, England, 1998
- [6] Ambisonics Explained: A Guide for Sound Engineers [En línea] Disponible en: <https://www.waves.com/ambisonics-explained-guide-for-sound-engineers>
- [7] The Ambisonics Toolkit [En línea] Disponible en: <http://www.ambisonictoolkit.net/documentation/reaper/>
- [8] FB360 Spatial Workstation user guide [En línea] Disponible en: <https://facebookincubator.github.io/facebook-360-spatial-workstation/Documentation/SpatialWorkstation/SpatialWorkstation.html>
- [9] Javascript mouse click coordenates for image [En línea] Disponible en: <https://stackoverflow.com/questions/34867066/javascript-mouse-click-coordinates-for-image>
- [10] [En línea] Disponible en: <https://www.w3schools.com/bootstrap/default.asp>
- [11] Bootstrap JavaScript Modal [En línea] Disponible en: <https://getbootstrap.com/docs/4.1/components/modal/>

5. ANEXOS

ANEXO A

En la primera carpeta Juego Colores se tienen los siguientes archivos:

- **Templ1_P1.rpp**: Primera plantilla donde se recogen las grabaciones de los 8 colores y se reproduce el audio a los altavoces.
- **Templ2_P1.rpp**: Segunda plantilla donde se añade el audio B-Format y se crea la mezcla para espacializarlo.
- **Juego Colores.mp4**: Video sin espacializar que se introduce en VR Virtual Player.
- **Colors_Game.wav**: Mezcla de audio espacial exportada de la segunda plantilla.
- **Colors_Game360.mp4**: Video 3D final con el video y audio 360°.

La segunda carpeta incluye los siguientes archivos:

- **Templ1_P2.rpp**: Primera plantilla con los 8 sonidos monofásicos de los elementos de la jungla.
- **Templ2_P2.rpp**: Segunda plantilla donde se añade el audio B-Format y se crea la mezcla para la espacialización binaural.
- **Templ3_P2.rpp**: Tercera plantilla donde se incorpora audio A-Format y se reproduce el sonido a la salida por 8 altavoces.
- **Juego Selva.mp4**: Video sin espacializar que se introduce en VR Virtual Player.
- **Jungle_Game.wav**: Mezcla de audio espacial exportada de la segunda plantilla.
- **Jungle_Game360.mp4**: Video 3D final con el video y audio 360°.

La tercera carpeta referente al Camino Fluvial contiene 10 subcarpetas con las diferentes ubicaciones. En cada una de ellas se encuentran los siguientes archivos:

- **TemplX_P3.rpp**: Plantilla que recoge los audios en Formato A y Formato B ambix, y crea la mezcla para la reproducción binaural.
- **Place_X.mp4**: Video sin espacializar que incluimos en el VR Virtual Player.
- **PlaceX.wav**: Mezcla de audio espacial exportada de la plantilla de Reaper con la ayuda del plugin FB360 Encoder.
- **PlaceX_360.mp4**: Video 3D final con video y audio 360°, listo para subir a la plataforma Youtube.

ANEXO B

Los sonidos característicos que se diferencian en cada una de las ubicaciones captadas:

1.- Puente de Landaben

- El sonido del tráfico
- Los pájaros
- Ciclista pasando por el puente

Audio: 001 (Ambisonics A) y 002 (Ambix)

2.- Presa de Harinera

- El sonido de la caída del agua por la presa
- Los pájaros de fondo

Audio: 003 (Ambisonics A) y 004 (Ambix)

3.- Presa de Harinera en lo alto

- Diferentes sonidos de pájaros
- El agua de la presa de fondo

Audio: 005 (Ambisonics A) y 006 (Ambix)

4.- Puente de Miluce

- El sonido del tráfico
- Los pájaros
- Sonido de un avión sobrevolando por encima en ese momento

Audio: 007 (Ambisonics A) y 008 (Ambix)

5.- Puente de San Jorge

- El sonido de los patos sobrevolando el puente en ese momento y los pájaros
- El sonido de la bajada del río
- El tráfico por encima del puente

Audio: 009 (Ambisonics A) y 010 (Ambix)

6.- Frontón Cuatro Vientos

- Sonido de ambulancia de fondo
- El pitido de un tren
- Los pájaros y el tráfico
- El sonido de un bebe llorando y paseantes hablando en el tercer y cuarto audio

Audio: 011 (Ambisonics A), 012 (Ambix), 013 (Ambix), 015 (Ambix)

7.- Puente de Santa Engracia

- Pisadas de un corredor
- El sonido de diferentes tipos de pájaros
- El sonido del tráfico (Autobús y moto)
- Bocinazo de un coche en el segundo audio

Audio: 016 (Ambisonics A) y 017 (Ambix)

8.- Puente Aranzadi

- El sonido del agua del río
- Las pisadas de un corredor y paseantes
- El ruido de la rueda de una bicicleta
- El ruido del tráfico
- Otro bocinazo de un coche

Audio: 018 (Ambisonics A) y 021 (Ambix)

9.- Puente de la Magdalena

- Sonido de los bastones de peregrinos
- El canto de los pájaros
- Sonido del tráfico y música proveniente de un coche

Audio: 001 (Ambisonics A) y 002 (Ambix)

10.- Las pasarelas

- El sonido de la bajada del río
- El ruido del lanzamiento de piedras en la orilla
- El chapoteo de un perro bañándose

Audio: 003 (Ambisonics A) y 006 (Ambix)